

BEST AVAILABLE COPY

WO 01/12659

PCT/IB00/01496

Query: 168 QPNGPPEAEPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQORNVAELRSISEQR 227
P P P SP + A + S S K P + P + + + +
Sbjct: 737 PPAPVSSPPPTPVSSPPALAP-VSSPPSVKSS--PPAPLSSPPAPQVKSSPPPVQVSS 793

Query: 228 PPQAPKSKKAPPPVARKPSVGVPPASPSYPRAEPLTAPP 268
PP APK SP P+A P V PP + P PL++PP
Sbjct: 794 PPPAPKSSP----PLA--P-VSSPPQVEKTSPPAPLSSPP 827

Score = 165 (24.8 bits), Expect = 6.0e-09, P = 6.0e-09
Identities = 79/264 (29%), Positives = 105/264 (39%)

Query: 5 PPPEEAFSSVASPEPAG-PSGSP--ELVSSPAASSSSATALQIQQPGSPDPPPP-APPAPA 60
PPP + + + P P G PS P +VS P S P GSP PP +PP PA
Sbjct: 517 PPPVK---TTSPPAPIGSPSPPPVSVVSPPPVKSPPPPA---PVGSPPPPEKSPPPPA 570

Query: 61 PASSAPGHVAKLPQKEPVGCSKG---GGPPREDVGAP---LVTPSLLQMVRLRSVGAPGG 114
P +S P V P V PP V +P + +P V AP
Sbjct: 571 PVASPPPPVKSPPPTLVASPPPPVKSPPPPAPVASPPPPVKSPPPTPVASPPPPAPVA 630

Query: 115 APTPALGPSAPQKPLRRALSGRASVPVAP---SSGLHAAVRLKACSLAASEGLSSAQPNG 171
+ P + P P+ SP P P S+ S+ +S + P
Sbjct: 631 SSPPPMKSPPPPTPVSSPPPEKSPPPPPPAKSTPPPEEYPTPTSVKSSPPPEKSLP-- 688

Query: 172 PPEAEPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQORNVAELRSISEQRPPQA 231
PP P PP T SK P SPE + + V+ + PP A
Sbjct: 689 PPTLIPSPPPQEKPTPSTPSKP-----PSSPEKSP-PKEPVSSPPQTPKSSPPPA 739

Query: 232 PKKSPKAPPPVARKPSVGV--PPASPSYPRAEPLTAPP 268
P SP P PV+ P++ PP+ S P PL++PP
Sbjct: 740 PVSSPP-PTPVSSPPALAPVSSPPSVKSSPPAPLSSPP 777

Score = 162 (24.3 bits), Expect = 1.3e-08, P = 1.3e-08
Identities = 76/272 (27%), Positives = 99/272 (36%)

Query: 2 ADFPPPEEAFSSVASPEPAG-PSGSP-ELVSSPAASSSSATALQIQQPGSPDPPPPAPPAPA 60
A P P SPEP PS P P + S A PP P P +PPA +
Sbjct: 427 ASAPMPSPHTPPDVSEPLEPEPSVPVAPAPMPMPTPHSPPADDYVPTPPVPGKSPATS 486

Query: 61 PASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGAPT-- 118
P+ A P V S PP+ VG+P P V+ S AP G+P+P
Sbjct: 487 PSPQVQPPAASTPPPSLVKLS----PPQAPVGSPP--PPP----VKTTSPAPIGSPSPPP 536

Query: 119 ---ALGPSAPQK-PLRRALSGRASVPVAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPE 174
+ P P K P A G SP P S A S + + PP
Sbjct: 537 PVSVVSPPPVKSPPPPAPVG--SPPPEKSPPPAPVASPPPPVKSPPPTLVASPPPP 594

Query: 175 AEPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQORNVAELRSISEQRPPQAPKK 234
+ PP +P ++ + P P A + + PP P+K
Sbjct: 595 VKSPPPPAPVASPPPPVKSPPPTPVASPPPPAPVASSPPPMKSPPPPTPVSSPPP-PEK 653

Query: 235 SPKAPPPVARKPSVGVPPASPSYPRAEPLTAPPTNGLP 273
SP PPP P PP P+ P + + PP LP
Sbjct: 654 SPPPPPPAKSTP----PPEEYPTPTSVKSSPPPEKSLP 688

Score = 159 (23.9 bits), Expect = 2.8e-08, P = 2.8e-08
Identities = 77/264 (29%), Positives = 103/264 (39%)

Query: 5 PPPEEAFSSVASPEPAGPSGSP-ELVSSPAASSSSATALQIQQPGSP--DPPAP---PAP 59
PPP V+SP P P SP P SS ++ PP +P PP P P P
Sbjct: 916 PPPA---MVSSP-PMTPKSSPP----PVVSSPPPTVKSSPPPPAPVSSPPATPKSSPPP 966

Query: 60 APASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGAPT 119
AP + P V P PV S P AP+ +P + V+ AP +P P
Sbjct: 967 APVNLPPPEVKSSPPPTPV-SPPPAKSSPPPPAPMSSPPPPPE-VKSPPPPAPVSSPPPP 1024

Query: 120 LGPSAPQKPLRRALSG-RASVPVAPSSGLHAAVRLKACSLAASEG---LSSAQPNGPPEA 175
+ P P+ ++ P PAP S V+ S + S P P +
Sbjct: 1025 VKSPPPPAPVSSPPPPVKSPPPPAPVSSPPPPVKSPPPPAPISSPPPPVKSPPPPAPVSS 1084

Query: 176 EPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQORNVAELRSISEQRPPQAPKKS 235
P P +SP A S ++ P P A + A ++ S PP AP S
Sbjct: 1085 PPPPVKSPPPPAPV----SSPPPIKSPPPP---APVSSPPPPAPVKPPS--LPPAPVSS 1135

Query: 236 PK--APPPVARKPSVGVPPPA-SPSYPRAEPLTAPP 268
P P +K +PPPA S P + PP
Sbjct: 1136 PPPVVTAPPKKEEQLPPPAESQPPPSFNDIILPP 1171

Score = 143 (21.5 bits), Expect = 1.8e-06, P = 1.8e-06
Identities = 59/179 (32%), Positives = 77/179 (43%)

Query: 3 DFPPEEAFSVASPEP-AGPSGSELVSSPAASSSSATA-LQIQQPGSP--DPPP---A 55
 + PPPE S P P + P +P+ PA SS ++ PP +P PPP +
 Sbjct: 970 NLPPPEVK--SSPPTPVSSPPAPKSSPPAPMSSPPPEVKSPPPAPVSSPPPPVKS 1027

Query: 56 PPAPAPASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGA 115
 PP PAP SS P V P PV PP + P S V+ AP +
 Sbjct: 1028 PFPAPVSSPPPPVKSPPPPAPVSSPP---PPVKSPPPPAPISSPPPPVKSPPPPAPVSS 1084

Query: 116 PTPALGPSAPQKPLRRALSG-RASVPAPSSGLHAAVRLKACSLAASEGLSSAQNGPPE 174
 P P + P P+ ++ P PAP S A +K SL +SS P PP
 Sbjct: 1085 PFPVKSPPPPAPVSSPPPIKSPPPAPVSSPPPPAP-VKPPSLPPAPVSS--P--PPV 1139

Query: 175 AEPRPPQ 181
 P PP+
 Sbjct: 1140 VTPAPPK 1146

Score = 133 (20.0 bits), Expect = 2.3e-05, P = 2.3e-05
 Identities = 50/132 (37%), Positives = 59/132 (44%)

Query: 1 MADFPPEEAFSVASPEPAGP-SGSELVSSP---AASSSSATALQIQQPGSP--DPPP 54
 M+ PPPE V SP P P S P V SP A SS ++ PP +P PPP
 Sbjct: 1001 MSSPPPE-----VKSPPPAPVSSPPPPVKSPPPPAPVSSPPPPVKSPPPPAPVSSPP 1055

Query: 55 ---APPAPAPASSAPGHVAKLPQKEPVGCSKG---GGPPREDVGAPLVTPSLLQMVRLRS 108
 +PP PAP SS P V P PV PP V +P P +
 Sbjct: 1056 PVKSPPPPAPISSPPPPVKSPPPPAPVSSPPPPVKSPPPPAPVSSP--PPPIKSPPPAP 1113

Query: 109 VGAPGGAPT--PALGPSAP 125
 V +P AP P+L P AP
 Sbjct: 1114 VSSPPAPVKKPSLPPAP 1132

Score = 110 (16.5 bits), Expect = 8.0e-03, P = 8.0e-03
 Identities = 41/121 (33%), Positives = 49/121 (40%)

Query: 5 PPPEEAFS----VASPEPAGP-SGSELVSSP---AASSSSATALQIQQPGSP--DPPP 54
 PPP S V SP P P S P V SP A SS ++ PP +P PPP
 Sbjct: 1060 PFPAPISSPPPPVKSPPPPAPVSSPPPPVKSPPPPAPVSSPPPPPIKSPPPAPVSSPP 1119

Query: 55 AP-----PAPAPASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPSLLQMVRLRS 108
 AP P PAP SS P V P K+ + PP E P +L +
 Sbjct: 1120 APVKPSLPPAPVSSPPVTPAPPKKE---EQSLPPAESQPPPSFNDIILPPIMANK 1176

Query: 109 VGAP 112
 +P
 Sbjct: 1177 YASP 1180

Score = 108 (16.2 bits), Expect = 1.3e-02, P = 1.3e-02
 Identities = 46/155 (29%), Positives = 67/155 (43%)

Query: 114 GAPTPALGPSAPQKPLRRALSGRASVPAPSSGLHAAVR-LKACS-LAASEGLSSAQNG 171
 G PTP GP + P + A S +P+P+ + L S + A + P+
 Sbjct: 408 GYPTGGGGPSSVPVGKPAAS---APMPSHTPPDVSPPEPLPEPSPVPAPAPMPMPTPHS 464

Query: 172 PPEAEPRPPQSPASTASFIFSKGSRKLQLERVPSPETQ---ADLQRNLVAELRSISEQR 227
 PP + PP P S + S +Q +P + Q + + +
 Sbjct: 465 PPADDYVPPTPPVPGKSPPATSPSPQVQPPAASTPPPSLVKLSPPQAPVGSPPPPVKTS 524

Query: 228 PPOAPKKSPPKAPPPVARKPSVGVPPPASPSYPRAEPLTAPP 268
 PP AP SP PPPV SV PPP S P P+ +PP
 Sbjct: 525 PP-APIGSPSPPPV---SVVSPPPVKSPPPPAPVGSPP 560

Pedant information for DKFZphmcf1_1c23, frame 1

Report for DKFZphmcf1_1c23.1

[LENGTH] 311
 [MW] 31534.58
 [PI] 9.48
 [KW] All Alpha
 [KW] LOW_COMPLEXITY 38.59 %

SEQ MADFPPEEAFSVASPEPAGPSGSELVSSPAASSSSATALQIQQPGSPDPPAPAPAPA
 SEGXX
 PRD ccc

SEQ PASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGAPTAL
 SEG xxxxxx.....XXXXXXXXXXXX


```

PRD      cccccccccccccccccccccccccccccccccccccchhhhhhhhhhhcccccccccccccc
SEQ      GPSAPQKPLRRALSGRASPV PAPSGLHAAVRLKACSLAASEGLSSAQPNGPPEAEPRPP
SEG      xxxxx.....xxxxxxxxxxxxxxxx
PRD      cccccchhhhhhhhhccccccccchhhhhhhhhhhhhhhhhhhcccccccccccccccccc

SEQ      QSPASTASFIFSKGSRKLQLERPVS PETQADLQRNLVAELRSISEQRPPQAPKKS PKAPP
SEG      xxxxx.....xxxxxxxxxxxxxxxx
PRD      cccccceeeccccchhhhhccccccccchhhhhhhhhhhhhhhhhhhcccccccccccccccccc

SEQ      PVARKPSVGVP PPASPSYPRAEPLTAPPTNGLPHTQDRTKRELAENG GVLQLVGPPEEKM
SEG      xxxxxxxxxxxxxxxxxxxxxxxxxxx.....
PRD      cccccccccccccccccccccccccccccccccccccchhhhhhhcccccecccccccccc

SEQ      LPGSDSQKELA
SEG      .....
PRD      ccccccccccc

```

(No Prosite data available for DKFZphmcf1_1c23.1)

(No Pfam data available for DKFZphmcf1_1c23.1)

DKFZphmcf1_1e15

group: transmembrane protein

DKFZphmcf1_1e15 encodes a novel 454 amino acid protein with similarity to C. elegans proteins and transporter proteins.

The novel protein is similar to the PTR2 family of proton/oligopeptide symporter proteins and the D-xylose-proton symporter. Thus, the protein is a transporter of a so far unknown compound.

The new protein can find application as a new transporter in eukaryotic cells, e.g. in drug transport into cells.

similarity to D-XYLOSE TRANSPORTER
membrane regions: 9

complete cDNA, complete cds, EST hits
matches cDNA encoding cell growth inhibiting factor (E12646)

Sequenced by DKFZ

Locus: unknown

Insert length: 1957 bp
Poly A stretch at pos. 1947, polyadenylation signal at pos. 1929

```

1  GGTGCAGCGC CCGGGCTGAG CGACAGCAAG TGCAGCGGGC TCCTACCCCG
51  GGTGAGGGGT GGCCTCCGCG TGGGATCGTG CCCTCTTCAG CCCGCTCCTG
101 TCCCCGACAT CACGTGTATT CCGCACGTCC CCTCCGCGCT GTGTGTCTAC
151 TGAGACGGGG AGGCGTGACA GGGCCCGGGT CCCTTCTCAG TGGTGTCTCTG
201 TGCTTCAGGG CAAGCTCCCC GTCTCCGGGC GCACTTCCCT CGCCTGTGTT
251 CGGTCCATCC TCCTTTCTCC AGCCTCCTCC CCTCGCAGGT GGGATCGTCG
301 GTGGGACCGG AGCGCGGGCG GCGCGGGCCC CCGGGACCA TGGCCGGGTC
351 CGACACCGCG CCCTTCTCTCA GCCAGGCGGA TGACCGGAC GACGGGCCAG
401 TGCCCTGGCAC CCCGGGGTTG CCAGGGTCCA CGGGGAACCC GAAGTCCGAG
451 GAGCCCGAGG TCCCGGACCA GGAGGGGCTG CAGCGCATCA CCGGCCTGTC
501 TCTCATGTGC TCGGCTCTCA TAGTGGCGGT GCTGTGCTAC ATCAATCTCC
551 TGAACATACAT GGACCGCTTC ACCGTGGCTG TGTTCATCTC CAGTTACATG
601 GTGTTGGCAC CTGTGTTTGG CTACCTGGGT GACAGGTACA ATCGGAAGTA
651 TCTCATGTGC GGGGGCATTG CCTTCTGGTC CCTGGTGACA CTGGGGTCAT
701 CCTTCATCCC CGGAGAGCAT TTCTGGCTGC TCCTCCTGAC CCGGGGCTG
751 GTGGGGGTGCG GGGAGGCCAG TTATTCCACC ATCGCGCCCA CTCTCATTGC
801 CGACCTCTTT GTGGCCGACC AGCGGAGCCG GATGCTCAGC ATCTTCTACT
851 TTGCCATTCC GGTGGGCACT GGTCTGGGCT ACATTGCAGG CTCCAAAGTG
901 AAGGATATGG CTGGAGACTG GCACTGGGCT CTGAGGGTGA CACCGGGTCT
951 AGGAGTGGTG GCCGTTCTGC TGCTGTTCTT GGTAGTGCGG GAGCCGCCAA
1001 GGGGAGCCGT GGAGCGCCAC TCAGATTGTC CACCCCTGAA CCCCACCTCG
1051 TGGTGGGCAG ATCTGAGGGC TCTGGCAAGA AATCTCATCT TTGGAATCAT
1101 CACCTGCTCTG ACCGGAGTCC TGGGTGTGGG CCTGGGTGTG GAGATCAGCC
1151 GCCGGCTCCG CCACTCCAAC CCCCAGGGCTG ATCCCTGGT CTGTGCCACT
1201 GGCCTCCTGG GCTCTGCACC CTTCTCTTTC CTGTCCCTTG CCGCGCCCG
1251 TGGTAGCATC GTGGCCACTT ATATTTTCAT CTTCAATTGGA GAGACCTCC
1301 TGTCCATGAA CTGGGCCATC GTGGCCGACA TTCTGCTGTA CGTGGTGATC
1351 CCTACCCGAC GCTCCACCGC CGAGGCCTTC CAGATCGTGC TGTCCACCT
1401 GCTGGGTGAT GCTGGGAGCC CTAACCTCAT TGGCCTGATC TCTGACCGCC
1451 TGCGCCGGAA CTGGCCCCCT TCCTTCTTGT CCGAGTTCGG GGCTCTGCAG
1501 TTCTCGCTCA TGCTCTGCGC GTTTGTTGGG GCACTGGGCG GCGCAGCCTT
1551 CCTGGGCACC GCCATCTTCA TTGAGGCCGA CCGCCGGCGG GCACAGCTGC
1601 ACGTGACAGG CCTGTGCAC GAAGCAGGGT CCACAGACGA CCGGATTGTG
1651 GTGCCCCAGC GGGGCCGCTC CACCCGCGTG CCCGTGGCCA GTGTGCTCAT
1701 CTGAGAGGCT GCCGCTCACC TACCTGCACA TCTGCCACAG CTGGCCCTGG
1751 GCCCACCCCA CGAAGGGCCT GGGCCTAACC CCTTGGCCTG GCCCAGCTTC
1801 CAGAGGGACC CTGGGCCGTG TGCCAGCTCC CAGACACTAC ATGGGTAGCT
1851 CAGGGGAGGA GGTGGGGGTC CAGGAGGGGG ATCCCTCTCC ACAGGGGCAG
1901 CCCCAGGGC TCGGTGCTAT TTGTAACGGA ATAAAATTG TAGCCAGAAA
1951 AAAAAA

```

BLAST Results

Entry E12646 from database EMBL:
cDNA encoding cell growth inhibiting factor.
Score = 3046, P = 2.2e-131, identities = 640/659

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 340 bp to 1701 bp; peptide length: 454
 Category: similarity to known protein

```

1 MAGSDTAPFL SQADDPDDGP VPGTPGLPGS TGNPKSEEP VPDQEGQLQRI
51 TGLSPGRSAL IVAVLCYINL LNYMDRETVA VFISYVMVLA PVFGYLGDRY
101 NRKYL MCGGI AFWSLVTLGS SFIPGEHFWL LLLTRGLVGV GEASYSTIAP
151 TLIADLFVAD QRSRMLSIFY FAIPVGSGLG YIAGSKVKDM AGDWHWALRV
201 TPGLGVVAVL LFLVVRPEP RGAVERHSDL PPLNPTSWWA DLRLARNLI
251 FGLITCLTGV LGVGLGVEIS RRLRHSNPR DPLVCATGLL GSAPFLFLSL
301 ACARGSIVAT YIFIFIGETL LSMNWAIVAD ILLYVVIPT RSTAEAFQIV
351 LSHLLGDAGS PYLIGLISDR LRRNWPPSFL SEFRALQFSL MLCAFVGALG
401 GAAFLGTAIF IEADRRRAQL HVQGLLHEAG STDDRIVVPQ RGRSTRVPVA
451 SVLI

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphmcf1_1e15, frame 1

TREMBL:CEC13C4_1 gene: "C13C4.5"; Caenorhabditis elegans cosmid C13C4,
 N = 3, Score = 441, P = 5.2e-76

TREMBL:CEC39E9_10 gene: "C39E9.10"; Caenorhabditis elegans cosmid
 C39E9, N = 2, Score = 449, P = 8.2e-69

TREMBL:CEF09A5_1 gene: "F09A5.1"; Caenorhabditis elegans cosmid F09A5,
 N = 3, Score = 413, P = 9.1e-60

TREMBL:ATF6H11_18 gene: "F6H11.180"; product: "predicted protein";
 Arabidopsis thaliana DNA chromosome 5, BAC clone F6H11 (ESSAII
 project), N = 3, Score = 193, P = 2.5e-24

SWISSPROT:XYLT_LACBR D-XYLOSE-PROTON SYMPORT (D-XYLOSE TRANSPORTER)., N
 = 1, Score = 180, P = 7.9e-11

>TREMBL:CEC39E9_10 gene: "C39E9.10"; Caenorhabditis elegans cosmid C39E9
 Length = 488

HSPs:

Score = 449 (67.4 bits), Expect = 8.2e-69, Sum P(2) = 8.2e-69
 Identities = 88/204 (43%), Positives = 125/204 (61%)

Query: 58 SALIVAVLCYINLLNYMDRFTVAVFISSYMLAPVFGYLGDRYNRKYL MCGGIAFWSLVT 117
 + ++ V Y N + + + VF+ S+MV +PV GYLGDR+NRK++M G+ W
 Sbjct: 29 AGVLTQVQTYYNISDSLGGLIQTVFLISFMVFSVPCGYLGDRFNKRWIMIIGVGIWLGAV 88

Query: 118 LGSSFIPGEHFWL LLLTRGLVGVGEASYSTIAPT L IADLFVADQRSRMLSIFYFAIPVGS 177
 LGSSF+P HFWL L+ R VG+GEASYS +AP+LI+D+F +RS + IFYFAIPVGS
 Sbjct: 89 LGSSFVPANHEWLFVLRSFVGIGEASYSNVAPSLISDMFNGQKRSTVFMIFYFAIPVGS 148

Query: 178 GLGYIAGSKVKDMAGDWHWALRVTPGLGVVAVLLFLVVRPEPRGAVER---HSDLPPL 233
 GLG+I GS V + G W W +RV+ G++ ++ L L EP RGA ++ D+
 Sbjct: 149 GLGFIVGSNVATLTGHWQWQIRVSAIAGLIVMIALVLFTYEPERGAADKAMGESKDVVVT 208

Query: 234 NPTSWWADLRALARNLIFGLITCLTG 259
 T++ DL L + L+ C G
 Sbjct: 209 TNTTYLEDLVILLKTPT--LVACTWG 232

Score = 267 (40.1 bits), Expect = 8.2e-69, Sum P(2) = 8.2e-69
 Identities = 74/212 (34%), Positives = 113/212 (53%)

Query: 249 LIFGLITCLTGV LGVGLGVEISRRL-----RHSNPRADPLVCATGLLGSAPFLFLSL 300
 L FG IT G++GV G +S+ L R RA PLV G L +APFL + +
 Sbjct: 277 LYFGAITTAGGLIGVIFGSM LSKWL VAGWGPFRRLQTDRAQPLVAGGGALLAAPFL LIGM 336

Query: 301 ACARGSIVATYIFIFIGETLLSMNWAIVADILLYVVIPTRRSTAEAFQIVLSHLLGDAGS 360

S+V YI IF G T + NW + D+L V+ P RRSTA ++ +++SHL GDA
 Sbjct: 337 IFGDKSLVLLYIMIFFGITFCFNWGLNIDMLTTVIHPNRRSTAFSYFVLVSHLFGDASG 396

Query: 361 PYLIGLISDRLRN--WPPSFLSEFRALQFSLMLCAFGALGGAFLGTAIFIEADRR-- 416
 PYLIGLISD +R +P ++ +L + C + L + +++ + +DR+

Sbjct: 397 PYLIGLISDAIRHGSTYPKD---QYHSLVSATYCCVALLLSAGLYFVSSLTLVSDRKKF 453

Query: 417 RAQLHVQGLLHEA--GSTD--DRIVVPQGRSTRV 447
 RA++ + L + STD +RI + S+R+

Sbjct: 454 RAEMGLDDLQSKPIRTSTDRLERIGINDDVASSRL 488

Score = 70 (10.5 bits), Expect = 5.9e-24, Sum P(2) = 5.9e-24
 Identities = 25/89 (28%), Positives = 41/89 (46%)

Query: 62 VAVLCYINLLNYMDRFTVAVFISSYMLAPVFGYLGDRYNRKYLMCGGIAFWSLVT--LG 119
 V L +NLLNY+DR+TVA ++ + LG +L+ +S V LG

Sbjct: 11 VTALFVVNLLNYVDRTVAGVLTQVQTYYNISDSLGLIQTVFLI--SFMVFSFVCGYL 68

Query: 120 SSFIPGEHFVLLLLTRGLVGVGEASYSTIAP 150
 F W++++ G + +G S+ P

Sbjct: 69 DRF---NRKWIMIIGVG-IWLGAVLGSSFVP 95

Pedant information for DKFZphmcfl_1e15, frame 1

Report for DKFZphmcfl_1e15.1

[LENGTH] 454
 [MW] 49013.35
 [PI] 7.66
 [HOMOL] TREMBL:CEC13C4_1 gene: "C13C4.5"; Caenorhabditis elegans cosmid C13C4 2e-51

[BLOCKS] BL01022D
 [PROSITE] MYRISTYL 11
 [PROSITE] CAMP_PHOSPHO_SITE 1
 [PROSITE] CK2_PHOSPHO_SITE 3
 [PROSITE] PROKAR_LIPOPROTEIN 1
 [PROSITE] GLYCOSAMINOGLYCAN 1
 [PROSITE] PKC_PHOSPHO_SITE 4
 [KW] TRANSMEMBRANE 8
 [KW] LOW_COMPLEXITY 15.42 %

SEQ MAGSDTAPFLSQADDDPGVPVPGTGLPGSTGNPKSEEEVDPQEGLRITGLSPGRSAL
 SEGXX
 PRD cchhh
 MEMMM

SEQ IVAVLCYINLLNYMDRFTVAVFISSYMLAPVFGYLGDRYNRKYLMCGGIAFWSLVTLGS
 SEGXX
 PRD hhhhhhhhhcc
 MEM MMMMMM.....MM

SEQ SFIPGEHFVLLLLTRGLVGVGEASYSTIAPTLIADLFVADQSRMLSIIFYFAIPVGSGLG
 SEGXX
 PRD cccccchhhhhhhhhcc
 MEMMM

SEQ YIAGSKVKDMAGDWHWALRVTPGLGVVAVLLFLVVRPPRGAVERHSDLPPLNPTSWWA
 SEGXX
 PRD eeeccchh
 MEM MMMMMMMM.....MM

SEQ DLRALARNLIFGLITCLTGLVGLGVEISRRLRHSNPRADPLVCATGLLGSAFPLFLSL
 SEGXX
 PRD hhhhhhhhhhhhhheeecc
 MEMMM

SEQ ACARGSIIVATYIFIGETLLSMNWAIVADILLYVVIPTRRSTAEAFQIVLSHLLGDAGS
 SEGXX
 PRD cccccchhhhhheeecc
 MEM MMMM.....MM

SEQ PYLIGLISDRLRNWPSPFLSEFRALQFSLMLCAFGALGGAFLGTAIFIEADRRRAQL
 SEGXX
 PRD ceeehhhhhhhhhcc
 MEM MMMMMMMM.....MM

SEQ HVQGLLHEAGSTDDRIVVPQGRSTRVPVASVLI

```

SEG .....
PRD hhhhhhhhhcccccccccccccccccccccccccc
MEM MMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

```

Prosite for DKFZphmcf1_1e15.1

PS00002	177->181	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	340->344	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	270->273	PKC_PHOSPHO_SITE	PDOC00005
PS00005	339->342	PKC_PHOSPHO_SITE	PDOC00005
PS00005	368->371	PKC_PHOSPHO_SITE	PDOC00005
PS00005	444->447	PKC_PHOSPHO_SITE	PDOC00005
PS00006	11->15	CK2_PHOSPHO_SITE	PDOC00006
PS00006	342->346	CK2_PHOSPHO_SITE	PDOC00006
PS00006	431->435	CK2_PHOSPHO_SITE	PDOC00006
PS00008	26->32	MYRISTYL	PDOC00008
PS00008	32->38	MYRISTYL	PDOC00008
PS00008	52->58	MYRISTYL	PDOC00008
PS00008	139->145	MYRISTYL	PDOC00008
PS00008	176->182	MYRISTYL	PDOC00008
PS00008	252->258	MYRISTYL	PDOC00008
PS00008	262->268	MYRISTYL	PDOC00008
PS00008	266->272	MYRISTYL	PDOC00008
PS00008	288->294	MYRISTYL	PDOC00008
PS00008	305->311	MYRISTYL	PDOC00008
PS00008	397->403	MYRISTYL	PDOC00008
PS00013	292->303	PROKAR_LIPOPROTEIN	PDOC00013

(No Pfam data available for DKFZphmcf1_1e15.1)

DKFZphmcfl_1g13

group: mammary carcinoma derived

DKFZphmcfl_1g13 encodes a novel 573 amino acid protein with very weak similarity to the human KIAA0543 protein and Musca domestica hermes transposase.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of mammary carcinoma-specific genes.

similarity to KIAA0766

complete cDNA, complete cds, few EST hits
on genomic level encoded by AC005020, no splicing, genomic?

Sequenced by DKFZ

Locus: unknown

Insert length: 2210 bp

Poly A stretch at pos. 2200, polyadenylation signal at pos. 2176

```

1  GAAACCTGAT CTCATAAAAC CTAGGTCACA AAGGACAGCC CTGCAAAACA
51  GACCCCTATTT GGATCAAGTG AGCCAGTTCC TGGAACTGA ATAATGACTC
101 CTGAATCAAG GGATACTACA GATTTGTCTC CAGGGGTAC CCAGGAGATG
151 GAAGGCATCG TGATAGTGAA GGTGGAGGAG GAAGATGAAG AAGACCATTT
201 TCAAAAGGAA AGAAACAAAG TAGAGTCATC GCCACAAGTT CTCAGTCGCT
251 CTACAACATAT GAATGAGAGA GCCTTATTGT CATCGTATT AGTTGCATAT
301 AGAGTGGCAA AAGAGAAAAT GGCTCACACA GCGGCTGAAA AAATTATCCT
351 TCCAGCATGT ATGGACATGG TACGGACAAT TTTTGATGAC AAATCAGCTG
401 ATAACTAAG AACTATACCT CTTAGTGATA ATACAATATC TCGTCGAATC
451 TGTACGATTG CAAAACATTT GGAAGCAATG CTTATTACAC GGCTGCAGTC
501 CGGTATAGAC TTTGCAATCC AACTCGATGA GAGCACTGAT ATTGCAAGTT
551 GTCCCACTCT CTTGGTTTAT GTCAGATATG TGTGGCAAGA TGATTTTGTA
601 GAGGATCTCT TATGTTGTTT AAATTTAAAT TCACATATA CTGGATTAGA
651 TTTATTTACT GAATTAGAAA ACTGCCTTCT TGGTCAGTAT AAATTAACT
701 GGAACATTG TAAAGGAATT TCAAGTGATG GAACAGCAAA TATGACCGGA
751 AAACACAGCA GACTTACTGA AAAATTGTTA GAAGCAACCC ACAACAATGC
801 TGTTTGGAAT CACTGTTTTA TTCATCGAGA AGCTTTGTA TCCAAAGAAA
851 TTTCAACAAG TCTGATGGAT GTATTGAAAA ATGCACTGAA AACTGTTAAT
901 TTTATTAAAG GAAGCTCACT GAATAGCCGA CTTCTCGAAA TATTTTGTTT
951 AGAGATTGGA GTGAACCACA CCCACTTATT GTTTCATACA GAAGTTCGTT
1001 GGCTTTCTCA AGGAAAAGTA TTGAGCAGAG TATATGAAT CAGGAACGAG
1051 ATTTACATTT TTCTCGTTGA AAAGCAATCT CATTGGCAA ATATTTTGA
1101 AGACGACATT TGGGTAACAA AATTGGCATA TTTAAGTGAT ATTTTGGCA
1151 TTCTTAATGA ATTAAGCCTG AAAATGCAGG GGAACCAAA TGATATATTT
1201 CAGTATCTTG AACATATTCT AGGATTCCAA AAGACGTTAT TATTGTGGCA
1251 AAGCAAGACTT AAAAGTAACC GCCCTAGCTA CTATATGTTT CCAACATTAT
1301 TGCAACACAT CGAAGAGAAC ATTATTAATG AAGACTGCTT AAAAGAAAAA
1351 AAATTAGAGA TATTGTTGCA TCTCACTTCT TTGCTCAAA CTTTAAATTA
1401 TTACTTTCCG GAAGAGAAAT TTGAATCATT AAAGGAAAAT ATTTGGATGA
1451 AAGATCCATT TGCTTTTCAA AACCAGAAT CAATAATTGA GTTAACTTG
1501 GAGCCTGAAG AAGAGAATGA ATTATTGCAG CTCAGTTCAT CATTACACT
1551 AAAGAATTAT TATAAGATAT TAAGTTTATC AGCATTTTGG ATTAAGATTA
1601 AAGATGACTT TCCACTGCTA AGTAGGAAGA GTATATTGCT GTTACTACCA
1651 TTCACAATA CATATTTGTG TGAAGTAGGA TTTTCAATCT TGACACGGTT
1701 AAAACAAAG AAGAGAAAATA GGCTCAATAG TGCACCAGAT ATGCGGGTAG
1751 CATTATCTTC ATGTGTTCCCT GACTGGAAGG AACTTATGAA CAGACAAGCA
1801 CACCCATCAC ATTAATACAA AACTTTACAA AATTCTGTGT ATAGCCAGGT
1851 GTGGTGGCTT ACGCCTGTAA TCCCAGCAGT GGGAGACCGA GGTGGGCAGA
1901 TCACCTGAGT TCAAGACCAG CCTGGCCAAC ATGGTGAAAC CCCATCTCTA
1951 CTAATAATAG AACCTTAGC CAGGCGTGGT GGCACATGCC TGCAGTCCCA
2001 GTTACTTGGG TGCCTGAGGC AGGAGAATCT CTTAAACCAG GAAGGCAGAG
2051 ATTGCAGTGA GCTGAGATAA TCCCAGTGCA TTCCAGCCTG GGCAACAGCG
2101 TGAGACTTCA TCTCAAAAAA AAAAATTGT ATTTGTACTT TTAAGGGAT
2151 TTTGCAGTAT GTTGTAGTTA AACGTTAATA AAATTATATT TGTAATTAGG
2201 AAAAAA

```

BLAST Results

Entry AC005020 from database EMBL:
Homo sapiens clone GS259H13; HTGS phase 1, 4 unordered pieces.
Score = 9110, P = 0.0e+00, identities = 1822/1822

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 94 bp to 1812 bp; peptide length: 573
 Category: similarity to unknown protein

```

1 MTPESRDSTD LSPGGTQEME GIVIVKVEEE DEEDHFQKER NKVESSPQVL
51 SRSTTMNERA LLSSYLWAYR VAKEKMAHTA AEKIILPACM DMVRTIFDDK
101 SADKLRTIPL SDNTISRRIC TIAKHLEAML ITRLQSGIDF AIQLDESTDI
151 ASCPTLLVYV RYVWQDDFVE DLLCCLNLNS HITGLDLFTE LENCLLGQYK
201 LNWKHCKGIS SDGTANMTGK HSRLTEKLE ATHNNAVWNH CFIHREALVS
251 KEISPSLMDV LKNAVKTVNF IKGSSLNSRL LEIFCSEIGV NHTHLLFTE
301 VRWLSQGVKL SRVYELRNEI YIFLVEKQSH LANIFEDDIW VTKLAYLSDI
351 FGILNELSLK MQGKNNDIFO YLEHILGFQK TLLWQARLK SNRPSYMF
401 TLLQHIEENI INEDCLKEIK LEILLHLTSL SQTFNYYFPE EKFSLEKNI
451 WMKDPFAFQN PESIIEELNLE PEEENELLQL SSSFTLNYY KILSLSAFWI
501 KIKDDFPLLS RKSILLLLPF TTYLCELGF SILTRLKTKK RNRLNSAPDM
551 RVALSSCVPD WKELMNRQAH PSH

```

BLASTP hits

Entry AC004877_3 from database TREMBLNEW:
 gene: "WUGSC:H_DJ0751H13.2"; product: "KIAA0543 protein"; Homo sapiens
 PAC clone DJ0751H13 from 7q35-qter, complete sequence.
 Score = 86, P = 4.4e-03, identities = 46/179, positives = 78/179

Entry MD36211_1 from database TREMBL:
 product: "Hermes transposase"; Musca domestica Hermes transposase
 gene, complete cds.
 Score = 105, P = 3.0e-02, identities = 101/465, positives = 202/465

Alert BLASTP hits for DKFZphmcf1_lg13, frame 1

TREMBL:AB018309_1 gene: "KIAA0766"; product: "KIAA0766 protein"; Homo
 sapiens mRNA for KIAA0766 protein, complete cds., N = 1, Score = 300, P
 = 1.1e-23

>TREMBL:AB018309_1 gene: "KIAA0766"; product: "KIAA0766 protein"; Homo
 sapiens mRNA for KIAA0766 protein, complete cds.
 Length = 607

HSPs:

Score = 300 (45.0 bits), Expect = 1.1e-23, P = 1.1e-23
 Identities = 120/485 (24%), Positives = 229/485 (47%)

```

Query:   89 CMD-MVRTIFDDKSADKLRTIPLSDNTISRRICTIAKHLEAMLITRLQSGIDFAIQLDES 147
          CM+ ++R + + L+ + LS + +RI +I ++L L R + +++ LD+
Sbjct:  124 CMEVLLREVLPEH-VSVLQGVDLSPDITRQILSIDRNLRNQLFNRRARDFKAYSLALDDQ 182

Query:   148 TDIASCPTLLVYVRYVWQD-DFVEDLLCCLNLNSHIT-GLDLFTELENCLLGQYKLNWKH 205
          +A LLV++R V + + EDLL +NL H + G + LE+ L L+ +
Sbjct:  183 AFVAYENYLLVFIRGVGPELEVQEDLLTIINLTHHFSVGALMSAILES--LQTAGLSLQR 240

Query:   206 CKGISSDGTANMTGKHSRLTEKLEATHNNAVWN--HC--FIHREALVSKEISPSLMDVL 261
          G+++ T M G++S L + E + WN H F+H E L S ++ + ++
Sbjct:  241 MVGLTTHTLRMIGENSGLVSYMREKAVSPNCWNVIHYSGLHLELLSSYDVDVN--QII 298

Query:   262 KNAVKTVNFIKSSSLNSRLLEIFCSEIGVNHHTHLLFTEVR-WLSQGVLSRVYELRNEI 320
          + + IK + + + +E H + + WL +GK L ++ LR E+
Sbjct:  299 NTISEWIVLIKTRGVRRPEFQTLTTESESEHGERVNGRCLNNWLRGKTLKLIFSRLKEM 358

Query:   321 YIFLVEKQSHLANIFEDDIWVTKLAYLSDFIGILNELSLKMQGKNNDIFQYLEHILGFQK 380
          FLV + + F D W+ +L DI L ELS +++ +HI F+
Sbjct:  359 EAFLVSVGATTVH-FSDKQWLCDFGLVDIMEHLRELSEELRVSKVFAAAAFDHICTFEV 417

```

Query: 381 TLLLWQARLKSNRPSYMFPTLLQHIEE---NIINEDCLKEIKLEILLHLTSLTSQTFNY 435
L L+Q ++ + FP L ++E N +E + +++ L + F

Sbjct: 418 KLNLFQRHIEEKNLTD--FPALREVVDLKKQNKEDKIFDPDRYQMWI--CRLQKEFER 473

Query: 437 YFPEEKFESLKENIWM-KDPFAFQNPESIIELNLEPEEENELLQLSSSFTLNKYYKILSL 495
+F + +F +K+++ + +PF F+ + I + +E L +L ++ L N Y+I L

Sbjct: 474 HFKDLRF--IKKDLLEFSNPFNFKEPYAPISVRVE-----LTKLQANTNLWNEYRIKDL 525

Query: 496 SAFWIKIK-DDFPLLSRKSIILLLPFTTTLCELGFSLTRKTKKRNRLNSA---PDMR 551
F+ + + +P++ + + F + +CE FS LTR + L R

Sbjct: 526 GQFYAGLSAESYPIIKGVACKVASLFDNSQICEKAFSYLTRNQHTLSQPLTDEHLQALFR 585

Query: 552 VALSSCVPDWKELMNRQAHPH 573
VA + P W +L+ R+ + S+

Sbjct: 586 VATTEMEPGWDDLVR-RERNESN 606

Score = 290 (43.5 bits), Expect = 1.5e-22, P = 1.5e-22
Identities = 120/485 (24%), Positives = 228/485 (47%)

Query: 89 CMD-MVRTIFDDKSADKLRTIPLSDNTISRRICTIAKHLEAMLITRLQSGIDFAIQLDES 147
CM+ ++R + + L+ + LS + +RI +I ++L L R + +++ LD+

Sbjct: 124 CMEVLLREVLPPEH-VSVLQGVLDSPDITRQRIILSIDRNLRLQFNRRDFKAYSIALDDQ 182

Query: 148 TDIASCPTLLVYRVYVWQD-DFVEDLLCCLNLNSHIT-GLDLFTELENCLLGQYKLNWKH 205
+A LLV++R V + + EDLL +NL H + G + Le+ L L+ +

Sbjct: 183 AFVAYENYLLVFIRGQPELEVQEDLLTIINLTHHFSVGALMSAILES--LQTAGLSLQR 240

Query: 206 CKGISSDGTANMTGKHSRLTEKLEATHNNNAVWNHCFIHREALVSKEISPSLMDV-LKNA 264
G+++ T M G++S L + E + WN IH + E+ S DV +

Sbjct: 241 MVGLTTHHTLRMIGENSGLVSYMREKAVSPNCWN--VIHSGFLHLELLSSY-DVDVNQI 297

Query: 265 VKTVN----FIKGSSLSNRLLEIFCSEIGVNHTHLLFHTEVR-WLSQGVLSRVYELRNE 319
+ T++ IK + + +E H + + WL +GK L ++ LR E

Sbjct: 298 INTISEWIVLIXTRGVRRPEFQTLTTESESEHGERVNGRCLNNWLRRGKTLKLIFSLRKE 357

Query: 320 IYIFLVEKQSHLANIFEDDIWVTKLAYLSDFIGILNELSLMKMQGNNDIFQYLEHILGQF 379
+ FLV + + F D W++ +L DI L ELS +++ +HI F+

Sbjct: 358 MEAFVLSVGATTVH-FSDKQWLCDFGLVDIMEHLRELSEELRVSKVFAAAAFDHICTFE 416

Query: 380 KTLLLWQARLKSNRPSYMFPTLLQHIEENIINEDCLKEIKL---EILLHLTSLTSQTFN 435
L L+Q ++ + FP L ++E ++ ++ K+ + + L + F

Sbjct: 417 VKLNLFQRHIEEKNLTD--FPALREVVDL--LKQNKEDKIFDPDRYQMWICRLQKEFE 472

Query: 436 YFPEEKFESLKENIWM-KDPFAFQNPESIIELNLEPEEENELLQLSSSFTLNKYYKILSL 494
+F + +F +K+++ + +PF F+ + I + +E L +L ++ L N Y+I

Sbjct: 473 RHFKDLRF--IKKDLLEFSNPFNFKEPYAPISVRVE-----LTKLQANTNLWNEYRIKD 524

Query: 495 LSAFWIKIK-DDFPLLSRKSIILLLPFTTTLCELGFSLTRKTKKRNRLNSA---PDM 550
L F+ + + +P++ + + F + +CE FS LTR + L

Sbjct: 525 LGQFYAGLSAESYPIIKGVACKVASLFDNSQICEKAFSYLTRNQHTLSQPLTDEHLQALFR 584

Query: 551 RVALSSCVPDWKELMNRQAHPH 573
RVA + P W +L+ R+ + S+

Sbjct: 585 RVATTEMEPGWDDLVR-RERNESN 606

Pedant information for DKFZphmcf1 lq13, frame 1

Report for DKFZphmcf1 1q13.1

```

[LENGTH]      573
[MW]           66276.85
[pI]           5.82
[HOMOL]        TREMBL:AB018309_1 gene: "KIAA0766"; product: "KIAA0766 protein"; Homo sapiens
mRNA for KIAA0766 protein, complete cds. ie-18
[PROSITE]      MYRISTYL      3
[PROSITE]      CK2_PHOSPHO_SITE      10
[PROSITE]      TYR_PHOSPHO_SITE      1
[PROSITE]      PKC_PHOSPHO_SITE      9
[PROSITE]      ASN_GLYCOSYLATION      2
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY      8.90 %

SEQ      MTPESRDTTDLSPGGTQEMEGIVIVKVEEDEDHDFQKERNKVESSPOVLRSRTTMNERA
SEG      .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
          ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ      LLSSYLVAIVRAKEKMAHTAAEKIILPACMDMVRTIFDDKSADKLRTIPLSDNTISRRIIC

```



```

SEQ      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhchhhhhhhhhhhccccceeecccchhhhhhh

SEQ      TIAKHLEAMLITRLQSGIDFAIQLDESTDIASCPTLLVYRVYWQDDFVEDLCLCNLNS
SEG      .....
PRD      hhhhhhhhhhhhhhhhhheeeccccccccccccccceeeeecchhhhhhhhhhhccce

SEQ      HITGLDLTELENCILGGYKLNWKHCKGISSDGTANMTGKHSRLTEKLEATHNNAVWNH
SEG      .....
PRD      eeehhhhhhhhhhhhhhhhhhccccccccccccccceeecccchhhhhhhhhhhcccccee

SEQ      CFIHREALVSKEISPSLMDVLKNVAKTVNFIKGSSLSNRLLLEIFCSEIGVNHTLLFHTE
SEG      .....
PRD      hhhhhhhhhhhhhccccchhhhhhhhhhhheeeccccchhhhhhhhhhhccccchhhhhhh

SEQ      VRWLSQGKVLRSVYELRNEIYIFLVEKQSHLANIFEDDIWTKLAYLSDIFGILNELSLK
SEG      .....
PRD      cccccccchhhhhhhhhhhhhhhhhhhhhhhchhhhhccccceehhhhhhhhhhhhhhhhh

SEQ      MQGKNNDIFYOLEHILGFQKTLLLWQARLKSNRPYSYMFPTLLQHIEENIINEDCLKIEK
SEG      .....xxxxx
PRD      hhccccccchhhhhhhhhhhhhhhhhhhhhhhccccccccchhhhhhhhhhhhhcchhhhh

SEQ      LEILLHLTSLSQTFNYYPPEEKFESLKENIWMKDPFAFQN PESIIENLPEEENELLQL
SEG      .....xxxxxxxxxxxxxxxxxxxxxx
PRD      hhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhccccccccceeecccchhhhhhhhh

SEQ      SSSF TLKNYYKILSLSAFWIKIKDDFP LLSRKSI LLLPFTTTYLC ELGFS ILTRLTKK
SEG      ....xxxxxxxxxxx.....
PRD      hhccccchhhhhhhhhhhhhccccccccchhhhhhhhhhhccccceehhhhhhhhhhhhh

SEQ      RNRLNSAPDMRVALSSCPVDKWEMLMNRQAHP SH
SEG      .....
PRD      hccccccccceeeccccccchhhhhhhcccccc

```

Prosites for DKFZphmcf1_1g13.1

PS000001	216->220	ASN_GLYCOSYLATION	PDOC000001
PS000001	291->295	ASN_GLYCOSYLATION	PDOC000001
PS000005	116->119	PKC_PHOSPHO_SITE	PDOC000005
PS000005	218->221	PKC_PHOSPHO_SITE	PDOC000005
PS000005	225->228	PKC_PHOSPHO_SITE	PDOC000005
PS000005	358->361	PKC_PHOSPHO_SITE	PDOC000005
PS000005	391->394	PKC_PHOSPHO_SITE	PDOC000005
PS000005	445->448	PKC_PHOSPHO_SITE	PDOC000005
PS000005	485->488	PKC_PHOSPHO_SITE	PDOC000005
PS000005	510->513	PKC_PHOSPHO_SITE	PDOC000005
PS000005	538->541	PKC_PHOSPHO_SITE	PDOC000005
PS000006	55->59	CK2_PHOSPHO_SITE	PDOC000006
PS000006	79->83	CK2_PHOSPHO_SITE	PDOC000006
PS000006	95->99	CK2_PHOSPHO_SITE	PDOC000006
PS000006	136->140	CK2_PHOSPHO_SITE	PDOC000006
PS000006	183->187	CK2_PHOSPHO_SITE	PDOC000006
PS000006	189->193	CK2_PHOSPHO_SITE	PDOC000006
PS000006	256->260	CK2_PHOSPHO_SITE	PDOC000006
PS000006	445->449	CK2_PHOSPHO_SITE	PDOC000006
PS000006	463->467	CK2_PHOSPHO_SITE	PDOC000006
PS000006	546->550	CK2_PHOSPHO_SITE	PDOC000006
PS000007	364->372	TYR_PHOSPHO_SITE	PDOC000007
PS000008	137->143	MYRISTYL	PDOC000008
PS000008	273->279	MYRISTYL	PDOC000008
PS000008	289->295	MYRISTYL	PDOC000008

(No Pfam data available for DKF2phmcf1 1q13.1)

DKFZphtes3_14g5

group: testes derived

DKFZphtes3_14g5 encodes a novel 379 amino acid protein with strong similarity to murine cell growth regulating nucleolar protein LYAR.

The novel protein is very similar to murine Ly-1 antibody reactive clone protein (LYAR). It contains a ATP/GTP-binding site motif A (P-loop, interacts with one of the phosphate groups of a ATP/GTP nucleotide), but not the zinc finger motif and and nuclear localization signals of lyar.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

strong similarity to cell growth regulating nucleolar protein LYAR, of mouse

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 1503 bp

Poly A stretch at pos. 1467, polyadenylation signal at pos. 1440

```
1 CCCAGAGGTC CGACCTGGGA GGCTGGGGCT CAGAGAGCAA TGTTTGCTGT
51 CTTCCATTGG AGTGACTGAA TTTCTACATG ACGGCTTTTT GACAAGACTT
101 AAAACCTGTC TTGGATAGAG AATATTAGC CATTACCTA AAAATGGTAT
151 TTTTACATG CAATGCATGT GGTGAATCAG TGAAGAAAAT ACAAGTGGAA
201 AAGCATGTGT CTGTTTGAG AACTGTGAA TGCCTTTCTT GCATTGACTG
251 CGGTAAAGAT TTCTGGGGCG ATGACTATAA AAACCACGTG AAATGCATAA
301 GTGAAGATCA GAAGTATGGT GGCAAAGGCT ATGAAGGTAA AACCCACAAA
351 GGGGACATCA AACAGCAGGC GTGGATTGAG AAAATTAGTG AATTAATAAA
401 GAGACCCAAT GTCAGCCCA AAGTGAGAGA ACTTTAGAG CAAATTAGTG
451 CTTTTGACAA CGTTCCAGG AAAAAGGCAA AATTCAGAA TTGGATGAAG
501 AACAGTTTAA AAGTTCATAA TGAATCCATT CTGGACCAGG TGTGGAATAT
551 CTTTTCTGAA GCTTCCAACA GCGAACCAGT CAATAAGGAA CAGGATCAAC
601 GGCCACTCCA CCCAGTGGCA AATCCACATG CAGAAATCTC CACCAAGGTT
651 CCAGCCTCCA AAGTGAAAGA CGCCGTGGAA CAGCAAGGGG AGGTGAAGAA
701 GAATAAAGA GAAAGAAAGG AAGAACGGCA GAAGAAAGG AAAAGAGAAA
751 AGAAGAAGT AAAGTTAGAA AACACCAGG AAAACTCAAG GAATCAGAAG
801 CCTAAGAAGC GCAAAAAGGG ACAGGAGGCT GACCTTGAGG CTGGTGGGGA
851 GGAAGTCCCT GAGGCCAATG GCTCTGCAGG GAAGAGGAGC AAGAAGAAGA
901 AGCAGCGCAA GGACAGCGCC AGTGAGGAAG AGGCACGCGT GGGCGCAGGG
951 AAGAGGAAGC GGAGGCACTC GGAAGTTGAA ACAGATTCTA AGAAGAAAAA
1001 GATGAAGCTC CCAGAGCATC CTGAGGGCGG AGAACCAGAA GACGATGAGG
1051 CTCTGCAAAA AGGTAAATTC AACTGGAAGG GAACTATTAA AGCAATTCTG
1101 AAACAGGCCC CAGACAATGA AATAACCATC AAAAAGCTAA GGAAGAAAGT
1151 TTTAGCTCAG TACTACACAG TGACAGATGA GCATCACAGA TCCGAAGAGG
1201 AACTCCTGGT CATCTTTAAC AAGAAAATCA GCAAGAACCC TACCTTTAAG
1251 TTATTTAAAG ACAAAAGTCAA GCTTGTGAAA TGAACATTG TGTATTTAAA
1301 AATTGAATCC ATTCTGCTGA CTTCTTCCTT TCACTGCTGT TTATAAAATG
1351 TGTAATGAAT TCTAACAACT CAAATTTTGC TTTTGAAGC TGTATTTTAA
1401 AGTTAAGAAA ATATATTTT GGTATAACTT TTATGAGAAA AATAAAATAT
1451 ATTCTGGTCC AAACCTCAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
1501 AAA
```

BLAST Results

No BLAST result

Medline entries

93259460:

LYAR, a novel nucleolar protein with zinc finger DNA-binding motifs, is involved in cell growth regulation.

Peptide information for frame 3

ORF from 144 bp to 1280 bp; peptide length: 379
 Category: strong similarity to known protein
 Classification: Cell division
 Prosite motifs: ATP_GTP_A (60-68)

```

1 MVFFTCNACG ESVKKIQVEK HVSVCNCEC LSCIDCGKDF WGGDYKNHVK
51 CISEDQKYGG KGYEGKTHKG DIKQQAUIQK ISELIKRPNV SPKVRELLEQ
101 ISAFDNVPRK KAKFQNMWKN SLKVHNESIL DQVWNIFSEA SNSEPVNKEQ
151 DQRPLHPVAN PHAEISTKVP ASKVKDAVEQ QGEVKKNKRE RKEERQKKRK
201 REKKELKLEN HQENSRNQKP KKRKKGQEAD LEAGGEEVPE ANGSAKRSK
251 KKKQKQDSAS EEEARVGAGK RKRHSEVET DSKKKMKLP EHPEGGEPE
301 DEAPAKGKFN WKGTIKAILK QAPDNEITIK KLRKKVLAQY YTVTDEHRS
351 EEELLVIFNK KISKNPFTKL LKDKVKLVK

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_14g5, frame 3

PIR:A40683 cell growth regulating nucleolar protein LYAR - mouse, N = 1, Score = 1410, P = 2.7e-144

SWISSPROT:YQ58_CAEEL HYPOTHETICAL 28.5 KD PROTEIN C16C10.8 IN CHROMOSOME III., N = 1, Score = 381, P = 2.9e-35

TREMBL:AC003058_18 gene: "F27F23.18"; product: "putative RNA-binding protein"; Arabidopsis thaliana chromosome II BAC F27F23 genomic sequence, complete sequence., N = 3, Score = 139, P = 4e-15

PIR:S70049 nucleic acid-binding protein YCR087c-a - yeast (Saccharomyces cerevisiae), N = 1, Score = 164, P = 1.4e-11

>PIR:A40683 cell growth regulating nucleolar protein LYAR - mouse
 Length = 388

HSPs:

Score = 1410 (211.6 bits), Expect = 2.7e-144, P = 2.7e-144
 Identities = 275/388 (70%), Positives = 317/388 (81%)

```

Query:      1 MVFFTCNACGESVKKIQVEKHVSVCNCECLSCIDCGKDFWGGDYKNHVKCISEDQKYGG 60
             MVFFTCNACGESVKKIQVEK VS CRNCECLSCIDCGKDFWGGDYK+HVKCISE QKYGG
Sbjct:      1 MVFFTCNACGESVKKIQVEKQVSNCRNCECLSCIDCGKDFWGGDYKSHVKCISEGQKYGG 60

Query:      61 KGYEGKTHKGDIKQQAUIQKISELIKRPNVSPKVRELLEQISAFDNVPRKKAKFQNMWKN 120
             KGYE KTHKGD KQQAUIQKI+ELIK+PNVSPKVRELL+QISAFDNVP KKAKFQNMWKN
Sbjct:      61 KGYEAKTHKGDAAKQQAUIQKINELIKRPNVSPKVRELLQQAISAFDNVPIKKAKFQNMWKN 120

Query:      121 SLKVHNESILDQVWNIFSEASNSEPVNKEQDQRPLHPVANPHAEIS-TKVPASKVKDAVE 179
             SLKVH++S+L+QVW+IFSEAS+SE ++Q Q P H A PHAE+ TKVP++K E
Sbjct:      121 SLKVHSDSVLEQVWDIFSEASSSE---QDQQPPSH-TAKPHAEMPITKVPSAKTNGTTE 176

Query:      180 QGEVKKNKRRERKEERQKKRKKREKKELKLENHQENSRNQKPKRRKKGQEADLEAGGEEVP 239
             +Q E KKNKRERKEERQK RK+EKKELKLENHQEN R QKPKRRKK QEA EA GE+
Sbjct:      177 EQTEAKKNKRERKEERQKNRKKKEKKELKLENHQENLRGQKPKRRKNQEAHEAAGEDGA 236

Query:      240 EANG-----SAGKRSKKKKQKQDSASEEEA----RVGAGKRKR-RHSEVETDSKKKKM 287
             + +G G+ S++ R E+ A + AGKRKR +HS E+ KKKKM
Sbjct:      237 DSGGPPEKKKAQGGQASEEGADRNGGPGEDRAEQTKTAAGKRKRPKHSGAESGYKKKKM 296

Query:      288 KLPEHPEGGEPEDEAPAKGKFNWKGTIKAILKQAPDNEITIKLRKKVLAQYYTVTDEH 347
             KLPE PE GE +D EAP+KGKFNWKGTIKA+LKQAPDNEI++KKL+KKV+AQY+ V ++
Sbjct:      297 KLPEQPEGEAKDHEAPSKGKFNWKGTIKAVLKQAPDNEISVKKLKKKVIAQYHAVMNDT 356

Query:      348 HRSEELLVIFNKKISKNPFTKLLKDKVKLVK 379
             EEELL IFN+KIS+NPTFK+LKD+VKL+K
Sbjct:      357 SHHEELLAIIFNRKISRNPFTKVLKDRVKLLK 388

```

Pedant information for DKFZphtes3_14g5, frame 3

Report for DKFZphtes3_14g5.3

[LENGTH] 379
[MW] 43634.03
[pI] 9.59
[HOMOL] PIR:A40683 cell growth regulating nucleolar protein LYAR - mouse 1e-122
[FUNCAT] 04.99 other transcription activities [S. cerevisiae, YCR087c-a] 2e-11
[BLOCKS] BL00603D Thymidine kinase cellular-type proteins
[BLOCKS] BL00530C
[PROSITE] ATP_GTP_A 1
[KW] All_Alpha
[KW] LOW_COMPLEXITY 18.73 %

SEQ MVFFTCNACGESVKKIQVEKHVSVCRNCECLSCIDCGKDFWGGDYKNHVKCISEDQKYGG
SEG
PRD cccccccccccccchhhhhhhheeeccccceeeccccccccccccccccceeecccccccc
SEQ KGYEGKTHKGDIKQAWIQKISELIKRPNVSPKVRELLEQISAFDNVPRKKAKFQNMKN
SEG
PRD cccccccccchhhhhhhhhhhhhhhccchhhhhhhhhhhccccchhhhhhhhhhhc
SEQ SLKVHNESILDQVWNIFSEASNSEPVNKEQDQRPLHPVANPHAEISTKVPASKVKDAVEQ
SEG
PRD cccccchhhhhhhhhhhhhhhccchhhhhhhhhccccccccccccceeeccccchhhhhh
SEQ QGEVKKNKRRERKEERQKKRKREKKELKLENHQENSRNQPKKRKKGQADLEAGGEEVPE
SEG ...xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccchhhhhchhhhhccccccc
SEQ ANGSAKRSKKKKQKQDSASEEEARVGAGKRRKRHSEVETDSKKKKMKLPEHPEGGEPEP
SEG ..xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.....xxxxxxxxxxxxx
PRD cccccccchhhhhhhhhccchhhhhhhhhccccccccccccccccchhhhhccccccccccc
SEQ DEAPAKGKFNWKGTIKAILKQAPDNEITIKLRKKVLAQYYTVTDEHHRSEEEELLVIFNK
SEG xxxxxx.....
PRD cccccceeehhhhhhhhhhhhccccccccchhhhhhhhhhhhhccchhhhhhhhhhhhh
SEQ KISKNPTEKLLDKVKLVK
SEGxxxxxxxxxxxxx
PRD cccccchhhhhhhhhhhccc

Prosite for DKFZphtes3_14g5.3

PS00017 60->68 ATP_GTP_A PDOC00017

(No Pfam data available for DKFZphtes3_14g5.3)

DKFZphtes3_14h21

group: nucleic acid management

DKFZphtes3_14h21 encodes a novel 648 amino acid protein with strong similarity to mus musculus RNA helicase and several RNA-dependent ATPases from the DEAD box family.

RNA helicases comprise a large family of proteins that are involved in basic biological systems such as nuclear and mitochondrial splicing processes, RNA editing, rRNA processing, translation initiation, nuclear mRNA export, and mRNA degradation. RNA helicases are essential factors in cell development and differentiation, and some of them play a role in transcription and replication of viral single-stranded RNA genomes. The members of the largest subgroup, the DEAD and DEAH box proteins, exhibit a strong dependence of the unwinding activity on ATP hydrolysis. The novel protein contains a DEAD-box and a ATP/GTP-binding site motif A (P-loop) and is a new member of this subgroup.

The new protein can find application in modulating RNA metabolism and gene expression.

strong similarity to RNA helicases

start at Bp 33 matches Kozak consensus ACNATg

Sequenced by BMFZ

Locus: unknown

Insert length: 2200 bp

Poly A stretch at pos. 2166, polyadenylation signal at pos. 2140

```
1 CAACGACGTC GGACGCGCCC CTTCTTGGAA CAATGTCCTCA CCACGGAGGA
51 GCTCCCAAGG CCTCTACGTG GGTCTGTGCT AGTCGGCGAA GCTCGACAGT
101 GTCCCGAGCG CCAGAGAGGA GGCCGGCGGA GGAGTTGAAT CGAACAGGTC
151 CTGAGGGATA TAGTGTCCGC AGAGGTGGTC GCTGGAGAGG CACCTCTAGG
201 CCCC CGGAGG CCGTGGCCGC TGGTCACGAG GAACTGCCGC TGTGTTTTC
251 TTTGAAGAGC CACTTTGTTG GCGCGGTAAT CGGTCTGGT GGGTCAAAAA
301 TAAAGAAATAT ACAAAGTACA ACAAACACCA CAATCCAAAT AATACAAGAA
351 CAACCAGAAAT CATTAGTCAA AATTTTGGC AGCAAGGCAA TGCAAACGAA
401 AGCAAAGCA GTGATAGACA ATTTTGTAA AAAGCTAGAA GAAAATTACA
451 ATTCAGAATG CGGAATTGAT ACTGCATTCC AACCTTCTGT TGGAAAAGAT
501 GGAAGCACAG ATAACAATGT TGTTCAGGA GATCGGCCAT TGATAGATTG
551 GGATCAAAT AGAGAGGAAG GTTTGAAATG GCAAAAAACA AAGTGGGCAG
601 ATTTACCACC AATTAAGAAA AACTTTTATA AAGAGTCCAC TGCCACAAGT
651 GCCATGTCAA AAGTAGAAGC AGATAGTTGG AGGAAAGAAA ATTTAATAT
701 AACGTGGGAT GACTTGAAGG ATGGGGAGAA ACGACCTATC CCAATCCTA
751 CCGTGCACAT TGATGACGCC TTCAATGTT ATCCTGAGGT TATGGAAAAC
801 ATTAAGAAAG CAGGTTTTC AAGCCAAACA CCTATTCAGT CACAGGCATG
851 GCCCATTTG TTGCAAGGAA TAGATCTTAT AGGAGTAGCC CAGACTGGAA
901 CAGGAAAGAC ATTGTGTTAT TTAATGCCTG GATTTATTCA TCTGGTCCTT
951 CAACCCAGCC TTAAGGTCA AAGGAATAGA CCCGGCATGT TAGTTCTAAC
1001 TCCCACTCG GAATTAGCAC TTCAAGTAGA AGGAGAATGT TGCAATATT
1051 CATATAAAGG GCTTCGGAGT GTTTGTGTAT ATGGTGGTGG AAATAGAGAT
1101 GAACAAATAG AAGAGCTTAA AAAAGGTGTA GATATCATAA TTGCAACTCC
1151 CGGAAGATTG AATGATCTGC AAATGAGTAA CTTCTCAAT CTGAAGAATA
1201 TAACCTACTT GGTTTGTAGT GAAGCAGACA AGATGTTGGA CATGGGATTT
1251 GAACCCAGCA TAATGAAGAT TTTGTAGAT GTGCGCCCAG ATAGGCAGAC
1301 AGTTATGACC AGTGCTACAT GGCCTCATTC AGTTCATCGC CTCGCACAAT
1351 CTTATTTGAA AGAACCAATG ATTGTCTATG TTGGTACATT GGATCTAGTT
1401 GCTGTAAATT CAGTGAAGCA AAATATAATT GTAACCACCG AGGAAGAGAA
1451 ATGGAGTCAC ATGCAAACTT TTCTACAGAG TATGTCATCC ACAGACAAAG
1501 TCATTGTCTT CGTTTCTCGA AAAGCTGTTG CGGATCACTT ATCAAGTGAC
1551 CTAATACTTG GAAATATATC AGTAGAGTCT CTGCATGGAG ATAGAGAACA
1601 GAGAGATCGG GAGAAAGCAT TAGAGAACTT TAAAAACAGC AAAGTGAGAA
1651 TACTAATTGC AACTGATCTA GCCTCTAGAG GACTTGATGT CCATGACGTT
1701 ACACATGTCT ATAATTTTGA CTTTCCACGG AATATTGAAG AATACGTACA
1751 CCGAATAGGG CGCACGGGAA GAGCAGGGAG GACTGTTGTT TCCATTACAA
1801 CTTTGACTAG AAATGATTGG AGGGTTGCCT CTGAATTGAT TAATATTCTG
1851 GAAAGAGCAA ATCAGAGTAT TCCAGAGGAG CTTGTATCAA TGGCTGAGAG
1901 GTTTGAGGCA CATCAACGGA AAAGGGAAT GGAAGAAAAA ATGGAAGAAC
1951 CTCAGGAAG GCCCAAGAAG TTTCATTAAT GTCTTCTGTA CTAGTGGGGT
2001 AGAGAATTCA AGATTTTTTA GAAATATAGT AAGACAGAAG TATTGGACAT
2051 GTTGCCAGTA TGAAGAGACC GGACTGATTT GACTGATTCT TAAAAATAA
2101 GTGTTTGAAT ATATAGAATC CAGTGTTTA TACTTTCTTT AATAAAAAA
2151 GAAGTATTTA AACTTGAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 33 bp to 1976 bp; peptide length: 648
 Category: strong similarity to known protein
 Classification: Nucleic acid management
 Prosite motifs: ATP GTP A (286-294)
 DEAD_ATP_HELICASE (394-403)

```

1 MSHHGGAPKA STWVVASRRS STVSRAPERR PAEELNRTGP EGYSVGRGGR
51 WRGTSRPPPEA VAAGHEELPL CFALKSHFVG AVIGRGGSKI KNIQSTTNTT
101 IQIIQEQEPES LVKIFGSKAM QTKAKAVIDN FVKKLEENYN SECGRIDTAFO
151 PSVGKDGSTD NNVVAGDRPL IDWDQIREEG LKWQKTKWAD LPPIKKNFYK
201 ESTATSAMSK VEADSWRKEN FNITWDDLKD GEKRPINPT CTFDADFQCY
251 PEVMENIKKA GFQKPTPIQS QAWPIVLQGI DLIGVAQTGT GKTLCYLMGP
301 FIHLVLQPSL KGQRNRPGLM VLTPTRELAL QVEGECKYS YKGLRSVCVY
351 GGGNRDEQIE ELKKGVDIII ATPGRNLNDLQ MSNFVNLKNI TYLVLDEADK
401 MLDMGFEQPI MKILLDVRPD RQTVMTSATW PHSVHRLAQS YLKEPMIVYV
451 GTLDLVAVSS VKQNIIVTTE EEKWSHMQTF LQSMSSTDKV IVFVSRKAVA
501 DHLSSDLILG NISVESLHGD REQDRREKAL ENFKTGKVRI LIATDLASRG
551 LDVHDVTHVY NFDFFRNIEE YVHRIGRTGR AGRTGVSITT LTRNDWRVAS
601 ELINILERAN QSIPEELVSM AERFEAHQRK REMERKMERP QGRPKKFH

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_14h21, frame 3

TREMBL:CEY54G11A_9 gene: "Y54G11A.3"; *Caenorhabditis elegans* cosmid
 Y54G11A, N = 1, Score = 1008, P = 1.1e-101

TREMBL:SPBP8B7_16 gene: "dbp2"; "SPBP8B7.16c"; product: "p68-like
 protein."; *S.pombe* chromosome II pl p8B7., N = 1, Score = 971, P =
 9.1e-98

PIR:S13757 RNA helicase DBP2 - yeast (*Saccharomyces cerevisiae*), N = 1,
 Score = 970, P = 1.2e-97

PIR:S14048 RNA helicase dbp2 - fission yeast (*Schizosaccharomyces*
pombe), N = 1, Score = 961, P = 1e-96

PIR:AS7514 RNA helicase HEL117 - rat, N = 2, Score = 888, P = 7.8e-91

>TREMBL:CEY54G11A_9 gene: "Y54G11A.3"; *Caenorhabditis elegans* cosmid
 Y54G11A
 Length = 504

HSPs:

Score = 1008 (151.2 bits), Expect = 1.1e-101, P = 1.1e-101
 Identities = 211/473 (44%), Positives = 298/473 (63%)

```

Query: 174 DQIREEGLKQKTKWADLPPIKKNFYKESTATSAMSKVEADSWRKENFNITWDDLKDGEK 233
      D++++E W K PI ++ YK +S + + ++
Sbjct: 23 DRLKDENFSWMK-----PIVRDLYKIPNEQKNLSPEQLQELYTNGGVMKVYPFREEST 75

Query: 234 RPIPNPTCTFDADFQCYPEVMENIKKAGFQKPTPIQSQAQWPIVLQGI DLIGVAQTGTGKT 293
      IP P +F+ AF +M I+K GF+KP+PIQSQ WP++L G D IGV+QTG+GKT
Sbjct: 76 VKIPPPVNSFEQAFGSNASIMGEIRKNGFEKPSPIQSQAQWPIVLQGI DLIGVAQTGTGKT 135

Query: 294 LCYLMPGFIHLVLQPSL-----KGQRNRPGLMVLTPPTRELALQVEGECKYSYKGLRSVC 348
      L +L+P ++ Q + + Q+ P +LVL+PTRELA Q+EGE KYSY G +SVC
Sbjct: 136 LAFLLPALHIDAQLAQYEKNDEEQKPSFVLVLSPTRELAQQIEGEVKKYSYNGYKSV 195

Query: 349 VYGGGNRDEQIEELKKGVDIIATPGRNLNDLQMSNFVNLKNITYLVLDEADKMLDMGFEP 408

```

+YGGG+R EQ+E + GV+I+IATPGRL DL ++L ++TY+VLDEAD+MLDMGFE
 Sbjct: 196 LYGGGSRPQVEACRGVEIVIATPGRLTDLSDNGVISLASVTYVVLDEADRLDMGFEV 255
 Query: 409 QIMKILLDVRPDRQTVMTSATWPHSVHRLAQSYLKEPMIVYVGTLDLVAVSSVKQNIIVT 468
 I +IL ++RPDR +TSATWP V +L Y KE ++ G+LDL + SV Q
 Sbjct: 256 AIRRILFEIRPDLVALTSATWPEGVRKLTDKYTKEAVMAVNGSLDLTSCSVTQFFFEV 315
 Query: 469 TEEKKW---SHMQTFLQSMSSD-KVIVFVSRKAVADHLSSDLILGNISVESLHGDREQR 524
 + ++ + FL + + K+I+V K +ADHLSSD + I+ + LHG R Q
 Sbjct: 316 PHDSRFLRVCEIVNFLTAAHGQNYKMIIFVKSVMADHLSSDFCMKGINSQGLHGGRSQS 375
 Query: 525 DREKALENFKTGKVRILIATDLASRGLDVHDVTHVYNFDFPRNIEEYVHRIGRTGRAGRT 584
 DRE +L ++G+V+IL+ATDLASRG+DV D+THV N+DFP +IEEYVHR+GRTGRAGR
 Sbjct: 376 DREMSLNMLRSGEVQILVATDLASRGIDVPDITHVLNDFPMDIEEYVHRVGRGTGRAGRK 435
 Query: 585 GVSITTLTRNDWRVASELINILERANQSIPEELVSMARFEAHQKREMERKMERPOGRP 644
 G +++ L ND LI ILE++ Q +P++L AE++ K + R RP R
 Sbjct: 436 GEAMSLFWNDRSNFEGLIQILEKSEQEVPDQLRRDAEKYRL---KCQSGRDGPRPSFRN 492
 Query: 645 KK 646
 K
 Sbjct: 493 NK 494

Pedant information for DKFZphtes3_14h21, frame 3

Report for DKFZphtes3_14h21.3

[LENGTH] 648
 [MW] 72873.51
 [pI] 8.84
 [HOMOL] TREMBL:CEY54G11A_9 gene: "Y54G11A.3"; Caenorhabditis elegans cosmid Y54G11A 1e-101
 [FUNCAT] 04.01.04 rna processing [S. cerevisiae, YNL112w] 2e-97
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YNL112w] 2e-97
 [FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YPL119c] 4e-72
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YOR204w] 2e-70
 [FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae, YOR204w] 2e-70
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YBR237w] 1e-61
 [FUNCAT] 1 genome replication, transcription, recombination and repair [H. influenzae, HI0892] 2e-49
 [FUNCAT] j mrna translation and ribosome biogenesis [H. influenzae, HI0231 RNA] 1e-48
 [FUNCAT] 04.99 other transcription activities [S. cerevisiae, YDL160c] 9e-45
 [FUNCAT] 04.05.01.07 chromatin modification [S. cerevisiae, YMR290c] 3e-44
 [FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YJL033w] 2e-36
 [FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YOR046c] 7e-32
 [FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YDR194c] 2e-28
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YGL064c] 5e-10
 [FUNCAT] 11.10 cell death [S. cerevisiae, YMR190c] 2e-08
 [FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YMR190c] 2e-08
 [FUNCAT] r general function prediction [M. jannaschii, MJ1401] 1e-07
 [BLOCKS] BL00039D DEAD-box subfamily ATP-dependent helicases proteins
 [BLOCKS] BL00039C DEAD-box subfamily ATP-dependent helicases proteins
 [BLOCKS] BL00039B DEAD-box subfamily ATP-dependent helicases proteins
 [BLOCKS] BL00039A DEAD-box subfamily ATP-dependent helicases proteins
 [PIRKW] nucleus 4e-96
 [PIRKW] RNA binding 3e-87
 [PIRKW] DEAD box 5e-50
 [PIRKW] transmembrane protein 4e-27
 [PIRKW] DNA binding 3e-67
 [PIRKW] recF recombination pathway 3e-10
 [PIRKW] ATP 4e-96
 [PIRKW] purine nucleotide binding 5e-50
 [PIRKW] P-loop 4e-96
 [PIRKW] hydrolase 9e-45
 [PIRKW] protein biosynthesis 5e-50
 [PIRKW] ATP binding 1e-61
 [SUPFAM] WW repeat homology 8e-88
 [SUPFAM] DEAD/H box helicase homology 4e-96
 [SUPFAM] unassigned DEAD/H box helicases 7e-87
 [SUPFAM] ATP-dependent RNA helicase DBP1 4e-96
 [SUPFAM] ATP-dependent RNA helicase DHH1 2e-43
 [SUPFAM] recQ protein 3e-10
 [SUPFAM] Bloom's syndrome helicase 5e-07
 [SUPFAM] translation initiation factor eIF-4A 5e-50
 [SUPFAM] recQ helicase homology 3e-10
 [SUPFAM] tobacco ATP-dependent RNA helicase DB10 8e-88
 [PROSITE] DEAD_ATP_HELICASE 1

[PROSITE] ATP_GTP_A 1
 [PFAM] Helicases conserved C-terminal domain
 [PFAM] KH domain family of RNA binding proteins
 [PFAM] DEAD and DEAH box helicases
 [KW] Alpha_Beta
 [KW] LOW_COMPLEXITY 8.49 %

SEQ MSHHGAGPKASTWVVASRRSSTVSRAPEERRPAEELNRTGPEGYSVGRGGRWRGTSRPPEA
 SEGXXXXXXXXXXXXXXXXX.....
 PRD ccc

SEQ VAAGHEELPLCFALKSHFVGAVIGRGGSKIKNIQSTTNTTIQIQEPESLVKIFGSKAM
 SEGXXXXXXXXXXXXXXXXX.....
 PRD cccccccccchhhhhcc

SEQ QTKAKAVIDNFVKLEENYNSECGIDTAFQPSVGKDGSTDNVNVAGDRPLIDWDQIREEG
 SEG
 PRD hhhhhhhhhhhhhhhhhhhhhcc

SEQ LKWQTKWADLPPIKKNFYKESTATSAMSKEADSWRKENFNITWDDLKDGKRPINPT
 SEG
 PRD chhhhhhhccccccccccccccccccccchhhhhhhhhhhhhhhheeecccccccccccccccc

SEQ CTFDFAQCYPEVMENIKKAGFQKPTPIQSQAQPIVLQIDLIGVAQTGTGKTLCLYMPG
 SEG
 PRD cccccccccchhhhhhhhhhhcc

SEQ FIHLVLQPSLKGQRNRPGLVLTPTRELALQVEGECCKYSYKGLRSVCVYGGGRNDEQIE
 SEG
 PRD eeeeeccccccccccccccccccccchhhhhhhhhhhhhhhccccccccccccccccchhh

SEQ ELKKGVDIIATPGRNLQMSNFVNLKNITYLVLDEADKMLDMGFEPQIMKILLDVRPD
 SEG
 PRD hhhheeeeeccccccccchhhhhhhccccccccccccccccccccchhhhhhhhhhhhhccc

SEQ RQTVMTSATWPHSVHRLAQSYLKEPMIVYVGTLDLVAVSSVKQNIIVTTEEEKWSHMQTF
 SEG
 PRD ceeeeeeccccchhhhhhhhhhhhhheeeeeccccccccccccccccccccchhhhhchhhhhhhhh

SEQ LQSMSSTDKVIVFVSRKAVADHLSSDLILGNISVESLHGDREQRDREKALENFKTGKVRI
 SEG
 PRD hhhheccccccccccccchhhhhhhhhhhhhccccccccccccccccchhhhhhhhhhhhhccc

SEQ LIATDLASRGLDVHDVTHVYNFDFPRNIEEYVHRIGRTGRAGRTGVSITTLTRNDWRVAS
 SEGXXXXXXXXXXXXX.....
 PRD eehhhhhhhccchhh

SEQ ELINILERANQSIPEELVSMAERFEAHQRKREMERKMERPQGRPKKFH
 SEGXXXXXXXXXXXXX.....
 PRD hhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccccc

Prosite for DKFZphtes3_14h21.3

PS00017	286->294	ATP_GTP_A	PDOC00017
PS00039	394->403	DEAD_ATP_HELICASE	PDOC00039

Pfam for DKFZphtes3_14h21.3

HMM_NAME	DEAD and DEAH box helicases		
HMM	*gLpPWILRnIyeMGFEKPTPIQQQAIPiILEGRDVMACAQTGSGKTAAAF		
	P++++NI+++GF KPTPIQ+QA+PI+L+G D+++ AQTG+GKT+++		
Query	248	QCYPEVMENIKKAGFQKPTPIQSQAQPIVLQIDLIGVAQTGTGKTLCLY	296
HMM	LIPMLQHIDwdFPWqpPQd..PrALILAPTRELAMQIQEEcRkFgkHMng		
	L+P ++H+ +P +++ Q+ P +L+L+PTRELA+Q++ EC K+++ +		
Query	297	LMPGFIHLVLQP-SLKGQRNRPGLVLTPTRELALQVEGECCKYSYK-G-	343
HMM	IRImcIYGGtnMRdQMRmLeRgPpHIVATPGRLLDHIERgtldLDRIeM		
	+R++C+YGG N ++Q+++L++G+ +I+IATPGRLL D+ +++ ++L++I++		
Query	344	LRSVCVYGGGRNDEQIEELKKGVD-IIATPGRNLQMSNFVNLKNITY	392
HMM	LVNDEADRMLDMGFIDQIRrIMrqiPmpwNRQTMFSATMPdeIqELARr		
	LV+DEAD+MLDMGF++QI++I+ ++ ++RQT+M SAT+P ++ +LA		
Query	393	LVLDEADKMLDMGFEPQIMKILLDVR--PDRQTVMTSATWPHSVHRLAQ	440


```

HMM          FMRNPiRInId.MdElTtnEnIkQwYiyVerEMWKfdCLcrLle*
++++P + ++ D +++ *KQ +I+ E++K + +++
Query        441 YLKEPMIVYVGTLDLVAVS-SVKQNIIVTT-EEEKWSHMQTFLLQ 482

HMM_NAME     KH domain family of RNA binding proteins

HMM          *rIiIPedhMGMIIGKGGsNIRqIREEYgvrINIPdecCeDstdRIITit
+ + +++G++IG+GGS I++I++ ++++I I++E+ + + + I
Query        71 CFALKSHFVGAVIGRGGSKIKNIQSTTNTTIQIIQE-Q-P---ESLVKIF 115

HMM          G*
G
Query        116 G 116

HMM_NAME     Helicases conserved C-terminal domain

HMM          *EileeWLknl...GirvmYIHGdMpQeERdeIMddFnnGEynVLicTD
+ +++ L+ + +I+V ++HGD++Q++R+++++F++G+ ++LI+TD
Query        497 KAVADHLSSDLILGNISVESLHGDRDREKALENFKTGKVRILIATD 545

HMM          VggRGIDIPdVNVHINYDMPWNPEqYIQRIGRTgRIG*
+++RG+D+ DV HV+N+D+P+N+E Y++RIGRTGR+G
Query        546 LASRGLDVHDTVHVYNFDFPRNIEEYVHRIGRTGRAG 582

```

DKFZphtes3_14p14

group: testes derived

DKFZphtes3_14p14 encodes a novel 159 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, few EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 3969 bp

Poly A stretch at pos. 3948, polyadenylation signal at pos. 3927

```

1 GAAGCCCAGG CTCTCCTTAG TTGACTGTGT GTTAATCACC CAGCAATTTT
51 ATTACTCAAC AGCTCTCCAG AGTTGCACAT TACAGCTGGG GTAGAAATTG
101 GGTGCTGAAG GCCAGGCAGA GCATTGGGCT GTAGGGAGGC CGATCCTCCT
151 CGGGCTGTGT ACCGGCGGGT CTTTGTCTT AGACCTGGGG TTCTTGGCCT
201 CACGGATTCC AAGGAATGGA ACGTTGGGCC ATGCGTGTGA ACGAGCTCTA
251 TGTCGATGAC CCAGACAAGG ACAGCGGTGG CAAGATCGAC GTCAGTCTGA
301 ACATCAGTTT ACCCAATCTG CACTGCGAGT TGGTTGGGCT TGACATTGAG
351 GATGAGATGG GCAGGCACGA AGTGGGCCAC ATCGACAAC CCATGAAGAT
401 CCCGCTGAAC AATGGGGCAG GCTGCCGCTT CGAGGGGCAG TTCAGCATCA
451 ACAAGGTATG GAAGCCCTGC CTCAGCCCTT TCTACCTGCT CCCCTTTCCT
501 GCTGTCTCCC CGCTCCCTGG AAACCTGGTTG TGGAGGCACT CACTCGACCT
551 GACCCCTGACA CAGCCCCCAG CAAGCGAGGG TTCGTGTCCA GCTGCCTGGC
601 CGTTCTCTGT GAGAATCTGC ATGGGGGTCC AGGCTCCCTG GGGTTTTAAG
651 CCCCTGATGG CTGGTTCAGG AAGGAGCTAC TCTTCTCTCC AGTGAGGGGG
701 ACAATGATGA GAAGACCTGA GGATTGTCAG CCCCCAGCCC TGGGTTCAAG
751 TCCCAGCTCT ACCCCTTCTT GGCCCTTACA AGTCACTTGA CCCATCTTAG
801 GCTGAGGGTG TGATGGCGAT AATAGTATCA CGATACCACC CACTTCACAA
851 AGTTTGTGTG GGGATTAAAT GAGCTAATGC AGATTCATTC ATTCAGAAAA
901 ATTTTGAAT GGCACGTTCT GTGTTCCAGG GTCGGTGATA GGCTCTGGGG
951 CAGCGTTTCT GGGCTGGTGG GGCTCCCAT CTGGTAGAGG GAGACAGTCT
1001 ACAAAACCAGA AAGCATCAGG GATGCTAAGT GCAGTGATGA GGAATAAAGC
1051 CAAGGGGAGT GAGATGAGGT GGGCTTGAAA GTACCTTGTC CGCTCAGAAG
1101 GACCATTCAA GGTTCAGTGT TGTTTGTGCC TCAGAACCAG GAGCTTCAGA
1151 TCCTAAGTCA AGTGGGTGAA CGCAGTGCCC TTGGGAGGGC CGAGGCACCC
1201 GGTGGCAGCT GGCAGGGTTT TGCTCAGCAC GTGCCGGCCT TCCTCGAAGC
1251 TCGGTACTGT CACAGTGGAG CCTCTCAACA ACGCTGTGAG CGAGCACCAT
1301 TTGACAGGTT AGGATGCTGG GGCCAGAGA GGTAAAGTGT CTGCCCCGAG
1351 GTACACAGCA TATCTGCATG TCCCACAAC CTCCCTTCCA GCCCCAGCCA
1401 AACTGAGCCA CTGGCCACTC CTGGCTTCTC CTGTCCCTC CTGACGCCCT
1451 TGCTCAGAAC GCCCTTCTC CAGACCCTGA CACCTGAGCT GGGGTGCAA
1501 AGTCACTGGC CACATCCAGC CCAAAGATAA ATTTTGTTTG TCCAGTATAG
1551 CATTTAACTG CATCAGAAC AGTATGAAAA GACCAGGAAT CCAGATTTCT
1601 GGCTTTTAAA AGTCAGAGGC TCTCACTACA CTGGGTCCGT GTTCCCGCTA
1651 TGACAATGAC CTGGCACCAA TGGGCAGTGT TCCCCTTTAG AGAGGGTGTG
1701 TGCTGTCCCT TCCCACAGTC CCTGGCAGGC GGCTGGAAGG CCAGGCCCTG
1751 TCATCTGTCA AGCAGGGTGG ACTTCTTACG TGACAGTTCA GGGCTCCCTT
1801 AAGTGCTAAA GCAGAAGCTG CAAGGCTTTC TTAAGGTTTC GAGTGTGCT
1851 GGGAGAAATC TGCTGCATGT TGTGGGTAA AGGGAGTCTC TCACCAGCCC
1901 AGGCCCTCAG GAGGAGGAGA TACCAGGAGG CAGGGATGCT GGGGGTCTGT
1951 GTTCACTGGG GGCTCTCTCT GCCCATGAGC TGCCACACAG CACCTTTGCC
2001 ATGCCCGGTA ATTTGGATT TATGGTGGTT GTGATGGAAG GCCATTTGAG
2051 GGTTTTGAAC AGGGAGGCAA TGTAATCAGA TTTATGCCTT AGAAGTGGAC
2101 TATCCAAATAG GTTGCCACCA GCCACATAAG GCTATTTAAA TTAATTCAAA
2151 TTAATGTAC AATTGAGTCA CTCATTCTCA TCAACCACAT TTCAAGTGCT
2201 CAAAGCCACG TGCTGGCTAG GGGCCACAGC GTTAGACAGT GCAGAGAGAA
2251 AGCACTTCCA TCGCTGAGGA AAGTTCTGCT GGACCGCACA CCCTTAGAAG
2301 GATGGCTCTG GTGGCCGGGC GCGGTGGCTC AAACCTGTAA TCCCAGCACT
2351 TTGGGAGGCC GAGGTGGGTG GATCAGGAG TCAGGAGATC GAGACCATCC
2401 CCGCTAATCAT GGTGAAACCC TGCCCTTACT AAAAATACAA AAAAAACAA
2451 AATTAGCCGG GCGTGGTTGC GGGCACCTGT AGTCCCAGCT ACTCAGGAGG
2501 CTGAGGCCGG AGAATGGCAT GAACCCGGGA GGTGGAGCTT GCAGTGAGCC
2551 AAGATCGTAC CACTGCACTC CAGTCTGGGC GACAGAGTGA GACTCCATCT
2601 CAAAACAAAC AAAAAAGGA TGGGGCTGGG CTGGAGAGGG TGGCAGGCAG
2651 TGGTTGTGGC AGTGGAGCTG GGGAGATGTG GTCGGATTAG GGAGGTAGAA
2701 TCAATAAGAC TCAGTGAAGA ATCGGATGTG GGGGTAAGGG CACATGTGGA

```

```

2751 AGCAAAGAAA CCTTTGACGT CTTTGTCTTG ACAACCGGGT GGTCTGTGTT
2801 CTAGACATGG AAGCTTAGAA AAGCCTGGAG TCTGTGGGAA GTAGGTAGGG
2851 CTGGGCACATG GTCATTCCAC TCTGGTTTCC TTGGGGTTC CCATTAGGTG
2901 TCTACAGGGA GAGGTGAAAT TGGAAGTTGG AGGTGTGGAG AGTTCAGGAG
2951 AGGGTTCTGG ACCACAGATG TTGAGGTGGG AGTCATTAGT GAATAGATGA
3001 TGTGGAAGT CATGGGTCTT CAGAGTGGGG GCTCCTTAAG CCTCCAGGCC
3051 AGCAGCATCA GCATCACCTG GGAGATTGTT AGGAATGCAG ATTCTCAGGC
3101 CCCCCTAAGA CCCACCGACT CTGTGCTAGA ACAAGCGCCC CTCAGAGATT
3151 CTGATGCCAC TGAAGTTTGA GGAGCATTGG TTTAAGCAAG ATTACCTACG
3201 GAGAGGCTGT AGATCCGTGT TCTAACCTG GGTCCACAG ACACCCCAA
3251 GAAGAGCGGA TTGAATGCAA GAGATCTATG AAGTTGGATG GGGGAAAAAT
3301 TGACATCTTT ATTTTGTCTA AACTCGATCT AAAGTTTAGC ATTTCCATCT
3351 GCGATGAATG TAGGCCACAA ACCACAGTAG TATTAGCAGT GCCTGGGACC
3401 TCCTCAACAA CAGAAATTGC CGGTATTTAT AGCACGTTAC AGTTGTTGCA
3451 GATAATTTCC AGAGACTGTT TATATGCACC ACTGTTTAA AATTACGGTG
3501 ATTTGGCCAGG TGCAGTGGCT CACACCTGTA ATCCCAGCAC TTGGGAGGC
3551 CAAAGTGGGT GGATCACTTG AGGAGTTCAA GACCAGCCTG GTCAACATGT
3601 CAAAACCTG TATCTACAAA AAAATACAAA AGTTAACCAG GCCTATGCTT
3651 GTAGTCACAG CTACTCGGGA GGCCGAGGTG GGAGGGTCTT CTGAGCCCAG
3701 GGAGGTAGAG GCTTCAGTGA GCTGAGATCG CACCACCACA CTCCAGCCTG
3751 GGTGACAGAG TGAAACCTT AATCAATCAG TCAATAAAAA TTACAGTAAT
3801 TATTAGACCC ACCACTAGGT CATCTTATTT GATGCATCAG TAAAGCAGCA
3851 TATTCAAATG TGGATTTTAA AATATTTTAA TTAATTTTAA AATATCTCTT
3901 TACTTTGTAA TCCTATGCAT TTTACGCATT AAAACATTTT AAGCATTTAA
3951 AAAAAAATAA AAAAAAATAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 216 bp to 692 bp; peptide length: 159
 Category: putative protein
 Classification: no clue

```

1 MERWAMRVNE LYVDDPKDSD GKKIDVSLNI SLPNLHCELV GLDIQDEMGR
51 HEVGHIDNSM KIPLNNGAGC RFEGQFSINK VWKPCLSPFY LLPFFAVSPL
101 PGNWLWRHSL DLTLTQPPAS EGSCPAAWPF LLRIWMGVQA PWGFKPLMAG
151 SGRSYSSSLQ

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_14p14, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_14p14, frame 3

Report for DKFZphtes3_14p14.3

```

{LENGTH}      159
{MW}           17778.55
{pI}           5.74
{FUNCAT}      99 unclassified proteins      [S. cerevisiae, YAL042w] 5e-04
{KW}           Alpha_Beta

```

```

SEQ  MERWAMRVNELYVDDPKDSDGKKIDVSLNISLPNLHCELVGLDIQDEMGRHEVGHIDNSM
PRD  ccchhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ  KIPLNNGAGCRFEGQFSINKVWKPCLSPFYLLPFFAVSPLPGNWLWRHSLDLTLTQPPAS
PRD  eecccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

SEQ EGSCPAAWPFLRLRIWMGVQAPWGFKPLMAGSGRSYSSLQ
PRD cccccchhhhhhhhhhhcccccccccccccccccccc

(No Prosite data available for DKFZphtes3_14p14.3)

(No Pfam data available for DKFZphtes3_14p14.3)

DKFZphtes3_14p7

group: testes derived

DKFZphtes3_14p7 encodes a novel 702 amino acid protein with very weak similarity to kinesin associated protein KAP3.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

weak similarity to kinesin associated protein KAP3

complete cDNA, complete cds, few EST hits

Sequenced by BMF2

Locus: unknown

Insert length: 2497 bp

Poly A stretch at pos. 2424, polyadenylation signal at pos. 2400

```
1  GGAATCCAAA  GAAACAGTTA  TGATGGGGGA  CTCTATGGTG  AAAATAAATG
51  GGATTTATTT  AACAAAATCA  AATGCTATTT  GCCACTTAAA  GAGTCACCCA
101  CTTCAGCTAA  CTGATGATGG  AGGCTTCAGT  GAAATAAAGG  AGCAAGAAAT
151  GTTCAAAGGA  ACAACATCTT  TACCATCTCA  TCTCAAGAAT  GGAGGGGACC
201  AGGGGAAGAG  ACATGCGAGG  GCCTCATCAT  GCCCCAGTAG  CTCAGACCTG
251  AGCAGGCTGC  AAACCAAAGC  AGTCCCAAAA  GCTGACCTGC  AAGAAGAGGA
301  CGCAGAAATA  GAAGTAGACG  AAGTCTTTTG  GAATACAAGG  ATTGTACCGA
351  TTTTGCCTGA  ATTAGAAAAG  GAAGAAAACA  TTGAAACGGT  TTGTGCTGCT
401  TGCACACAAC  TTCATCATGC  TTTAGAGGAA  GGAACATGCT  TTGGAATAAA
451  ATTTAAGGGA  AGAAGTATTC  TCCTGAAGAC  CCTGTGTAAA  CTAGTTGATG
501  TTGGTTCAGA  CTCGCTCAGC  CTTAAACTTG  CAAAAATAAT  TCTAGCACTT
551  AAAGTGAGTA  GAAAGAATCT  TCTTAATGTC  TGCAAACTTA  TATTTAAAT
601  TAGCAGGAAT  GAGAAGAATG  ATTCTTTGAT  TCAAAATGAC  AGCATTCTGG
651  AATCATTATT  GGAGGTACTA  AGAAGTGAAG  ACCTGCAAC  TAACATGGAA
701  GCTTTTTTAT  ACTGTATGGG  GTCTATAAAG  TTCATTTCTG  GAAATCTGGG
751  ATTTCTTAAT  GAAATGATCA  GCAAAGGTGC  TGTGGAATA  CTGATAAATT
801  TGATAAAACA  AATAAATGAG  AACATCAAGA  AATGTGGTAC  ATTTTGCCT
851  AATTCGGGCC  ACTTGCTAGT  CCAGGTGACT  GCTACATTGA  GAAACTTGGT
901  TGATTCATCA  TTAGTAAGAA  GTAAGTTCC  AACATCAGT  GCCCTTCCCC
951  AGCTCTGCAC  GGCAATGGAA  CAGTACAAGG  GTGACAAGGA  CGTCTGTACC
1001  AATATTGCCA  GAATATTGAG  CAACTTACT  TCTTACCGTG  ACTGCTGCAC
1051  AGCCTTGGCC  AGCTATTCCA  GATGTTATGC  CTTATTTCTG  AATCTAATTA
1101  ACAAAATACCA  GAAGAAGCAG  GATTTAGTCG  TCCGTGTGT  TTTTATCTT
1151  GGCAACCTGA  CGGCAAAAA  TAACCAGGCT  CGTGAACAAT  TTTCCAAAGA
1201  GAAAGGGAGC  ATCCAAACTC  TGCTGTCATT  ATTCCAGACG  TTCCATCAGC
1251  TGGATCTGCA  TTCCCAAG  CCGGTGGGCC  AACGAGGCGA  GCAGCAGG
1301  GCGCAGAGGC  CGCCGTCAGA  GGCAGAGGAC  GTGCTCATCA  AGCTGACTCG
1351  TGTGCTGGCC  AACATTGCCA  TCCACCCGGG  CGTGGGCCCG  GTGCTGGCCG
1401  CCAACCCGGG  GATAGTGGGC  CTGCTCCTGA  CCACGCTGGA  ATACAAGTCA
1451  CTTGATGATT  GTGAGGAGCT  GGTGATCAAT  GCTACAGCGA  CAATCAACAA
1501  TTTATCTTAC  TACCAAGTGA  AGAATTCCAT  AATTCAGAC  AAAAAAGCTAT
1551  ATATTGCTGA  ATTGCTCTTA  AAGCTTCTTG  TCAGTAACAA  CATGGATGGA
1601  ATCCTGGAGG  CTGTGCGTGT  TTTCCGAAAT  CTCTCCAGG  ACCATGATGT
1651  CTGCGATTTC  ATTGTGCAGA  ACAATGTCCA  CAGGTTTATG  ATGGCGCTGC
1701  TGGATGCTCA  GCATCAGGAT  ATCTGCTTTT  CTGCTGTGG  TGTCTCTCTC
1751  AATCTCACTG  TGGATAAAGA  CAAGCGTGTC  ATCTTGAAG  AAGGAGGTGG
1801  CATTAAAAAG  TTAGTGGACT  GTTAAAGAGA  TTTGGGTCTT  ACTGATTGGC
1851  AGCTGGCCTG  CTGGTGTGT  AAAACTTTAT  GGAACCTCAG  TGAACCATC
1901  ACTAATGCTT  CGTCATGTTT  TGGAAATGAA  GACACCAACA  CACTCTTACT
1951  CTTGCTCTCA  TCATTTTGTG  ATGAAGAACT  AGCACTGGAT  GGCAGTTTGT
2001  ATCCAGACCT  AAAAACTAT  CACAACTCC  ATTGGGAAAC  AGAATTCAAA
2051  CCTGTGGCAC  AGCAGCTTCT  AAACCGAATT  CAGAGACATC  ACACCTTCCT
2101  GGAACCCCTG  CCCATTCCCT  CTTTCTAACA  TGATGCAGAT  TAACAGTAGA
2151  AACGAGAACT  CACGCTCTCC  TCATTCTTAA  GAACTGGTAA  CAAACGTGAA
2201  CATTTTTTTC  AGCAATTAACA  AATGTGGAAA  GTTTTCAAG  AACTGGTTTT
2251  AGTGAGTAGC  TGAAGTATTT  TTTAAATTA  AGCATTCTCT  CTGTGTTAGT
2301  ATTATGGAAT  AATGAATATA  CACATTATAT  TTCTGTGTA  GAGAAATGTA
2351  AGATGAAAT  ATGTGCATTT  TCAAGTAAAT  GACTTTTCT  TCTATTCTCT
2401  ATTAACAAT  TTAGTTCTAG  TCTTAAAAA  AAAAAA  AAAAAA
2451  AAAAAA  AAAAAA  AAAAAA  AAAAAA  AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 20 bp to 2125 bp; peptide length: 702

Category: putative protein

```
1 MMGDSMVKIN GIYLTksNAI CHLkshPLQL TDDGGFSEIK EQEMFKGTTS
51 LPSHLKNGGD QGKRharASS CPSSDLSRL QTKAVPKADL QEEDAEIEVD
101 EVFWNTRIVP ILRELEKEEN IETVCAACTQ LHHALEEGNM LGNKFGRSI
151 LLKTLCKLVD VGSDSLSLKL AKIILALKVS RKNLLNVCKL IFKISRNEKN
201 DSLIQNDSIL ESLLVLRSE DLQTNMEAFI YCMGSIKFIS GNLGFLNEMI
251 SKGAVEILIN LIKQINENIK KCGTFLPNSG HLLVQVTATL RNLVDSSLVR
301 SKFLNISALP QLCTAMEQYK GDKDVCTNIA RIFSKLTSYR DCCTALASYS
351 RCYALEFLNI NKYQKKQDLV VRVVFILGNL TAKNNQAREQ FSKEKGSIQT
401 LLSLFQTFHQ LDLSHQKPVG QRGEQHRARQ PPSEAEDVLI KLTRVLANIA
451 IHPGVGPVLA ANPGIVGLLL TTLEYKSLDD CEELVINATA TINNLSYYQV
501 KNSIIQDKKL YIAELLLKLL VSNNDGILE AVRVFGNLSQ DHVCDFIVQ
551 NNVHREMMAL LDAQHQDICF SACGVLLNLT VDKDKRVILK EGGGIKKLVD
601 CLRD LGPTDW QLACLVCKTL WNFSENITNA SSCFGNEDTN TLLLLSSFL
651 DEELALDGSF DPKLKNYHKL HWETEFKPVA QLLNRIQRH HTFLEPLPIP
701 SF
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_14p7, frame 2

TREMBL:MMD367_1 product: "KAP3B"; Mus musculus mRNA for KAP3B, complete cds., N = 2, Score = 97, P = 0.00039

>TREMBL:MMD367_1 product: "KAP3B"; Mus musculus mRNA for KAP3B, complete cds.

Length = 772

HSPs:

Score = 97 (14.6 bits), Expect = 3.9e-04, Sum P(2) = 3.9e-04
Identities = 45/163 (27%), Positives = 77/163 (47%)

```
Query: 442 LTRVLANIAIHPGVGPVLAANPGIVGLLLTTLEYKSLDDCEELVINATATINNLSYYQVK 501
      L +++ NI+ H G P          VG L + S D+ EE VI T+ NL+ +
Sbjct: 483 LMKMIRNISQHDG--PTKNLFIDYVGDLAAQI---SSDEEEFVIECLGTLANLTIPDL 537
```

```
Query: 502 -NSIIQDKKLYIAELLKLLVSNNDG-ILEAVRVFGNLSQDHDVCDFIVQNNVHREMMMA 559
      +++++ KL+ L KL          D +LE V + G +S D + ++ + ++
Sbjct: 538 WELVLKEYKL-VPFLKDKLKPGAEDDLVLEVIMIGTVSMDDSCAALLAKSGIIPALIE 596
```

```
Query: 560 LLDAQHQDICFSACGVLL---NLTVDKDKR-VILKEGGGIKKLVDCLRD 604
      LL+AQ +D F C ++ + + R VI+KE L+D + D
Sbjct: 597 LLNAQQEDDEF-VCQIIYVFYQMVFHQATRDVIIKETQAPAYLIDLMDH 644
```

Score = 77 (11.6 bits), Expect = 3.9e-04, Sum P(2) = 3.9e-04
Identities = 42/178 (23%), Positives = 82/178 (46%)

```
Query: 169 KLAKIILALKVSRKNLLNVCK-LIFKISRNEKNDSLIQNDSILESLLVLRSEDLQTNME 227
      K K L V ++ LL V L+ ++ + + + ++N +I+ L++ L + N E
Sbjct: 263 KTFKKYQGLVVKQEQLLRVALYLLNLAE DTRTELKMRNKNIVHMLVKALDRD----NFE 318
```

```
Query: 228 AFLYCMGSIKIFISGNLGLNEMISKGAVEILINLIKQINENIKKCGTFLPNSGHLLVQVT 287
      + + +K +S + N+M+ VE L+ +I +E++ L + +
Sbjct: 319 LLILVVSFLKKLSIFMENKNDMVEMDIVEKLVKMIPCEHEDL-----LNITLR 366
```

```
Query: 288 ATLRLNVDSSLVRSKFLNISALPQLCTAM--EQYKGDVDCT--NIARI--FSKLTSYRD 341
      L D+ L R+K + + LP+L + E YK +C +I+ F + +Y D
Sbjct: 367 LLLNLSFDOTGL-RNKMVQVGLLPKLTALLGNENYK-QIAMCVLYHISMDDRFKSMFAYTD 424
```

Query: 342 CCTAL 346

C L

Sbjct: 425 CIPQL 429

Score = 69 (10.4 bits), Expect = 2.6e+00, Sum P(2) = 9.2e-01
Identities = 35/146 (23%), Positives = 70/146 (47%)

Query: 512 IAELLKLLVSNMMDGILEAVRVFGNLSQDHDVCD FIVQNNVHRFMMALLDAQHQDICFS 571
I +L+K L +N + ++ V LS + + +V+ ++ ++ ++ +H+D+

Sbjct: 304 IVHMLVKALDRDNFELLILVVSFLKKLSIFMENKNDMVEMDIVEKLVKMIPEHEDELLNI 363

Query: 572 ACGVLLNLTVDKDKRVILKEGGGIKKLVDCLRDLGPTDW-QLACLVCCKTLWNFSENITNA 630
+LLNL+ D R + + G + KL L G ++ Q+A +C L++ S +

Sbjct: 364 TLRLLLNLSFDTGLRNKMVQVGLLPKLTALL---GNENYKQIA--MC-VLYHISMD-DRF 416

Query: 631 SSCFGNEDT-NTLLLLSSFLDEELALD 657

S F D L+ +L DE + L+

Sbjct: 417 KSMFAYTDCIPQLMKMLFECSDERIDLE 444

Score = 68 (10.2 bits), Expect = 3.2e-03, Sum P(2) = 3.2e-03
Identities = 18/58 (31%), Positives = 30/58 (51%)

Query: 190 LIFKISRNEKN-DSLIQNDSILESLEVLRS-----DLQTNMEAFLYCMGSIKFISG 241

LI +++RN N + L+ N++ L +L VLR + +L TN+ +C S G
Sbjct: 155 LILQLARNPDNLEELLNLTALGALARVLRDQKQSVELATNIIYIFFCFSSFSHFHG 212

Score = 65 (9.8 bits), Expect = 6.4e+00, Sum P(2) = 1.0e+00
Identities = 26/122 (21%), Positives = 53/122 (43%)

Query: 283 LVQVTATLRNL---VDSSLVRSKFLNISALPQLCTAMEQYKGDQDVCTNIARIFSKLTS 338

+++ TL NL +D LV ++ +P L ++ + D+ + I S
Sbjct: 521 VIECLGTLANLTIPDLWELVLKEY---KLVPFLKDKLPGAAEDDLVLEV-IMIGTVS 576

Query: 339 YRDCCTALASYSRCYALFLNLINKYQKKQDLVVRVVFILGNLTAKNNQAREQFSKEKSGI 398

D C AL + S + L+N Q+ + V +++++ + + R+ KE +
Sbjct: 577 MDDSCALLAKSGIIPALIELLNAQQEDDEFVCQIIYVFYQMV-FQATRDVIKETQAP 635

Query: 399 QTLSSL 404

L+ L

Sbjct: 636 AYLIDL 641

Score = 65 (9.8 bits), Expect = 6.4e+00, Sum P(2) = 1.0e+00
Identities = 44/177 (24%), Positives = 79/177 (44%)

Query: 481 CE-ELVINATATIN-NLSYYQ-VKNSIIQDKKLYIAELLKLLVSNMMDGILEAVRVFGN 537

CE E ++N T + NLS+ ++N ++Q + + L LL + N I A+ V +
Sbjct: 355 CEHEDELLNITLRLLNLNLSFDTGLRNKMVQ---VGLLPKLTALLGNENYKQI--AMCVLYH 409

Query: 538 LSQDHDVCD-FIVQNNVHRFMMALLDAQHQDICFSACGVLLNLTVDKDKRVILKEGGGIK 596

+S D F + + + M L + + I +NL +K ++ EG G+K
Sbjct: 410 ISMDDRFKSMFAYTDCIPQLMKMLFECSDERIDLELISFCINLAANKRNVQLICEGNGLK 469

Query: 597 KLVDCLRDLGPTDWQLACLVCCKTLWNFSENITNASSCFGNEDTNTLLLLSSFLDEELAL 656

L+ R L D L+ K + N S++ + F + L +SS +EE +
Sbjct: 470 MLMK--RALKLK----PLMKMIRNISQHDGPTKNLF-IDYVGDLAAQISSDEEEEFVI 522

Query: 657 D 657

+

Sbjct: 523 E 523

Score = 61 (9.2 bits), Expect = 1.6e-02, Sum P(2) = 1.6e-02
Identities = 20/66 (30%), Positives = 34/66 (51%)

Query: 304 LNISALPQLCTAM-EQYKGDQDVCTNIARIFSKLTSYRDCCTALASYSRCYALFLNLINK 362

LN +AL L + E +K ++ TNI IF +S+ + Y + AL +N+I+
Sbjct: 171 LNETALGALARVLRDQKQSVELATNIIYIFFCFSSFSHFGLITHY-KIGALCMNIIDH 229

Query: 363 YQKKQDL 369

K+ +L

Sbjct: 230 ELKRHEL 236

Pedant information for DKFZphtes3_14p7, frame 2

Report for DKFZphtes3_14p7.2

[LENGTH] 708
[MW] 79266.35
[pI] 6.57

```

[FUNCAT]      30.25 vacuolar and lysosomal organization      [S. cerevisiae, YEL013w] 3e-04
[FUNCAT]      06.04 protein targeting, sorting and translocation [S. cerevisiae, YEL013w]
3e-04
[FUNCAT]      09.25 vacuolar and lysosomal biogenesis      [S. cerevisiae, YEL013w] 3e-04
[BLOCKS]      BL00923F Aspartate and glutamate racemases proteins
[BLOCKS]      BL00288B Tissue inhibitors of metalloproteinases proteins
[PROSITE]     MYRISTYL          9
[PROSITE]     AMIDATION         1
[PROSITE]     CK2_PHOSPHO_SITE   12
[PROSITE]     PKC_PHOSPHO_SITE   7
[PROSITE]     ASN_GLYCOSYLATION 11
[KW]          Alpha_Beta
[KW]          LOW_COMPLEXITY     7.49 %

```

```

SEQ  ESKETVMGDSMVKINGIYLTKSNAICHLKSHPLQLTDDGGFSEIKEQEMFKGTTSLPSH
SEG  .....
PRD  ccccccecccccecccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ  LKNGGDQKGRHARASSCPSSDLSRLQTKAVPKADLQEEDAIEVDEVFWNTRIVPILRE
SEG  .....xxxxxxxxxxxxx.....
PRD  cccccccchhhhhccccccccchhhhhccccccccchhhhhhhhhhhccccccccceehhhhhh

```

```

SEQ  LEKEENIETVCAACTQLHHALEEGNMLGNKFKGRSILLKTLCKLVDVGSLSLKLAKII
SEG  .....xxxxxxxxxxxxx.....
PRD  hhhhhcchhhhhhhhhhhhhhhhhccccccccccccccccccccchhhhhheeeccccchhhhhhhh

```

```

SEQ  LALKVSRKNLLNVCKLIFKISRNEKNDSLIONDSILESLLVLRSEDLQTNMEAFLYCMG
SEG  xxxxx.....
PRD  hhhhhhhhhhhhhhhhhccccccccccccccccccccchhhhhhhhhhhccchhhhhhhhhhhcc

```

```

SEQ  SIKFISGNLGFLENMISKGAVEILINLIKQINENIKKCGTFLPSNGHLLVQVTATLRNLV
SEG  .....
PRD  ceeeeccccchhhhhhhcchhhhhhhhhhhhhccccccccccccceeeehhhhhhhhhhh

```

```

SEQ  DSSLVRSKFLNISALPQLCTAMEQYKGDQDVCTNIARIFSKLTSYROCCTALASYSRCYA
SEG  .....
PRD  ccchhhhhheeeccccchhhhhhhhhhhccccceeeehhhhhhhhhhhccchhhhhhhhhhhhh

```

```

SEQ  LFLNLINKYQKKQDLVVRVVFILGNLTAKNNQAREQFSKEKGSIQTLLSLFQTFHQDLHLH
SEG  .....
PRD  hhhhhhhhhhhhhhhheeeccccccccchhhhhhhhhhhhhchhhhhhhhhhhhhhhhhcc

```

```

SEQ  SQKPVGQRGEQHRAQRPPSEADVLIKLTRLVLANIAIHGPGVPLAANPGIVGLLLTLE
SEG  .....
PRD  cccccccccccccccccccccchhhhhhhhhhhhhhhccccccccceeeccccchhhhhhhh

```

```

SEQ  YKSLDDCEELVINATATINNLSYYQVKNSIIQDKKLYIAELLKLLVSNMMDGILEAVRV
SEG  .....xxxxxxxxxxxxx.....
PRD  hhccccchhhhhhhheeeccccccccceeeehhhhhhhhhhhhhhhhhccccchhhhhhhh

```

```

SEQ  FGNLSDQHDVCDFIVQNNVHREFMALLDAQHQDICFSACGVLLNLTVDKDKRVILKEGGG
SEG  .....
PRD  cccccccccceeeccccchhhhhhhhhhhccccceeeccccceeeccccceeecccccc

```

```

SEQ  IKKLVDCLRLDGPFDWLACLVCVCKTLWNFSEINITASSCFGNEDTNTLLLLSSFLDEEL
SEG  .....xxxxxxxxxxxxx.....
PRD  hhhhhhhhhccccccccchhhhhhhccccccccccccccccccccceeeehhhhhhhh

```

```

SEQ  ALDGSFDPDLKNYHKLHWETEFKPVAAQLLNRIQRHHTFLEPLPIPSF
SEG  xxx.....
PRD  hhccccccccchhhhhhhhhhhchhhhhhhhhhhhhhhheeecccccc

```

Prosite for DKFZphtes3_14p7.2

PS00001	206->210	ASN_GLYCOSYLATION	PDOC00001
PS00001	212->216	ASN_GLYCOSYLATION	PDOC00001
PS00001	311->315	ASN_GLYCOSYLATION	PDOC00001
PS00001	385->389	ASN_GLYCOSYLATION	PDOC00001
PS00001	493->497	ASN_GLYCOSYLATION	PDOC00001
PS00001	500->504	ASN_GLYCOSYLATION	PDOC00001
PS00001	543->547	ASN_GLYCOSYLATION	PDOC00001
PS00001	584->588	ASN_GLYCOSYLATION	PDOC00001
PS00001	628->632	ASN_GLYCOSYLATION	PDOC00001
PS00001	632->636	ASN_GLYCOSYLATION	PDOC00001
PS00001	635->639	ASN_GLYCOSYLATION	PDOC00001
PS00005	173->176	PKC_PHOSPHO_SITE	PDOC00005
PS00005	186->189	PKC_PHOSPHO_SITE	PDOC00005
PS00005	241->244	PKC_PHOSPHO_SITE	PDOC00005

PS00005	295->298	PKC_PHOSPHO_SITE	PDOC00005
PS00005	344->347	PKC_PHOSPHO_SITE	PDOC00005
PS00005	387->390	PKC_PHOSPHO_SITE	PDOC00005
PS00005	421->424	PKC_PHOSPHO_SITE	PDOC00005
PS00006	79->83	CK2_PHOSPHO_SITE	PDOC00006
PS00006	201->205	CK2_PHOSPHO_SITE	PDOC00006
PS00006	214->218	CK2_PHOSPHO_SITE	PDOC00006
PS00006	218->222	CK2_PHOSPHO_SITE	PDOC00006
PS00006	230->234	CK2_PHOSPHO_SITE	PDOC00006
PS00006	320->324	CK2_PHOSPHO_SITE	PDOC00006
PS00006	344->348	CK2_PHOSPHO_SITE	PDOC00006
PS00006	439->443	CK2_PHOSPHO_SITE	PDOC00006
PS00006	477->481	CK2_PHOSPHO_SITE	PDOC00006
PS00006	483->487	CK2_PHOSPHO_SITE	PDOC00006
PS00006	654->658	CK2_PHOSPHO_SITE	PDOC00006
PS00006	698->702	CK2_PHOSPHO_SITE	PDOC00006
PS00008	17->23	MYRISTYL	PDOC00008
PS00008	64->70	MYRISTYL	PDOC00008
PS00008	144->150	MYRISTYL	PDOC00008
PS00008	384->390	MYRISTYL	PDOC00008
PS00008	402->408	MYRISTYL	PDOC00008
PS00008	473->479	MYRISTYL	PDOC00008
PS00008	533->539	MYRISTYL	PDOC00008
PS00008	580->586	MYRISTYL	PDOC00008
PS00008	641->647	MYRISTYL	PDOC00008
PS00009	67->71	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3_14p7.2)

DKFZphtes3_15a13

group: testes derived

DKFZphtes3_15a13 encodes a novel 387 amino acid protein with weak similarity to *S.cerevisiae* Hop1.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to *S.cerevisiae* Hop1

complete cDNA, complete cds, potential start codon at Bp 116, 3 EST hits

S.cerevisiae Hop1p is a meiosis-specific protein

Sequenced by GBF

Locus: unknown

Insert length: 1848 bp

Poly A stretch at pos. 1766, no polyadenylation signal found

```
1  GGAAAGCGCA  TCGCGCTCGG  GCACAGCGCG  TGCAGCCTCG  TGCAGCTCTT
51  CTGGTCTCCG  GCGCCCGCCC  CTCAGACGTA  ATGTTGAATT  AAAGAAAATA
101  CTTTATCAGA  AGAAGATGGC  CACTGCCCAG  TTGCAGAGGA  CTCCCATGAG
151  TGCACCTGGT  TTTCCCAATA  AGATATCAAC  TGAACCCAG  TCTTTGGTGT
201  TAGTGAAGAG  GCTTCTAGCA  GTTTCAGTAT  CCTGTATCAC  GTATTTGAGG
251  GGAATATTCC  CAGAAATGCG  TTATGGAACA  AGATATCTAG  ATGATCTTTG
301  TGTCAAAAAT  CTGAGAGAAG  ATAAAAATTG  CCCAGGATCT  ACACAGTTAG
351  TGAATGGAT  GCTAGGATGT  TATGATGCTT  TACAGAAAAA  ATATGTATAC
401  ACAAACCCAG  AAGATCCTCA  GACAATTTCA  GAATGTTACC  AATTCAAATT
451  CAAATACACC  AATAATGGAC  CACTCATGGA  CTCATAAGT  AAAAACCAAA
501  GCAACGAATC  TAGCATGTTG  TCTACTGACA  CCAAGAAAGC  AAGCATCTCT
551  CTCATTGCGA  AGATTTATAT  CCTAATGCAA  AATCTGGGCG  CTTTACCTAA
601  TGATGTTTGT  TTGACCATGA  AACTTTTTTA  CTATGATGAA  GTTACACCCC
651  CAGATTACCA  GCCTCCCGGT  TTTAAGGATG  GTGATTGTGA  AGGAGTTATA
701  TTTGAAGGGG  AACCTATGTA  TTTAAATGTG  GGAGAAGTCT  CAACACCTTT
751  TCACATCTTC  AAAGTAAAAG  TGACCACTGA  GAGAGAACGA  ATGGAAAATA
801  TTGACTCAAC  TATACTATCA  CCAAAACAAA  TAAAAACACC  ATTTCAAAAA
851  ATCCTGAGGG  ACAAAGATGT  AGAAGATGAA  CAGGAGCATT  ATACAAGTGA
901  TGATTGGGAC  ATTGAAACTA  AAATGGAAGA  ACAGGAAAAA  AACCCTGCAT
951  CTTCTGAAC  TGAAGAACCA  AGTTTAGTTT  GTGAGGAAGA  TGAAATTATG
1001  AGGTCTAAAG  AAAGTCCAGA  TCTTTCTATT  TCTCATTCTC  AGGTTGAGCA
1051  GTTAGTCAAT  AAAACATCTG  AACTTGATAT  GTCTGAAAGC  AAAACAAGAA
1101  GTGGAAAAGT  CTTTCAGAA  AAAATGGCAA  ATGGAATCA  ACCAGTAAAA
1151  TCTTCCAAAG  AAAATCGGAA  GAGAAGTCAA  CATGAATCTG  GGAGAATAGT
1201  CCTCCATCAC  TTTGATCTT  CTAGTCAAGA  GTCAGTGCCA  AAAAGGAGAA
1251  AGTTTAGTGA  ACCAAAGGAA  CATATATAAA  AATTATTTTT  GTTCTGCAGG
1301  CTTGCAGAGT  TCTTCTCACC  ATTTAACTG  AAGGACCTTA  TATTATATTT
1351  CCCTAACTCT  GAAGATGTAT  ATGTAGTTTA  AAGCAGTTTG  TACACTAAAA
1401  CTAAGTTTTT  GGCTGACTGT  CATATTGTGG  TCCTTAATCT  TGAGATAAAT
1451  CCAATAGAAC  TTTTGAATAA  AAGCAAAAGT  ACAAATGTCA  TAATTGATTG
1501  GGTAAATAAGT  AAAATTTCAA  AATTGATTTT  GTTCATTACC  TACTTAATAT
1551  TTCCTTTAAA  TATATACTAA  CTGTTAAGGC  CCTCTAATGC  CATTTTCTTA
1601  AACAGTAATG  TTTACTTTGG  TATTAAATTT  TGGTATGGAT  TCACTTTTAA
1651  CTTATGTATA  AATTATACCA  TTAACTGGC  TCTTTTGTCA  TTGTGCTGTT
1701  ATTAATAACAA  TGTTCTTCAA  TATTTTGACA  TAATGTATTA  ACATTTTAAT
1751  ATATAATGTA  CAATTTAAAA  AAAAAAAAAA  AAAAAAAAAA  AAAAAAAAAAGG
1801  GGCGGCGGCT  CTAGAGGATC  CAAGCTTACG  TACAAAAAAA  AAAAAAGG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 116 bp to 1276 bp; peptide length: 387
Category: similarity to known protein

```

1 MATAQLQRTPT MSALVFPNKI STEHQSLVLV KRLLAHSVSC ITYLRGIFPE
51 CAYGTRYLDD LCVKILREDK NCPGSTQLVK WMLGCYDALQ KKYVYTNPED
101 PQTISECYQF KFKYTNNGPL MDFISKQSN ESSMLSTDTK KASILLIRKI
151 YILMQNLGPL PNDVCLTMKL FYYDEVTPPD YQPPGFKDGD CEGVIFEGEP
201 MYLNVGEVST PFHIFKVKVT TERERENID STILSPKQIK TPFQKILRDK
251 DVEDEQEHYT SDDLDIETKM EEQEKNPASS ELEEPSLVCE EDEIMRSKES
301 PDLISHSQV EQLVNKTSSEL DMSSEKTRSG KVFQNMANG NQPVKSSKEN
351 RKRSQHESGR IVLHHFDSSS QESVPKRRKF SEPKEHI

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_15a13, frame 2

TREMBL:ATAC2130_3 product: "F1N21.3"; The sequence of BAC F1N21 from Arabidopsis thaliana chromosome 1, complete sequence., N = 1, Score = 274, P = 5.7e-22

TREMBL:SC9877_9 gene: "hop1"; S.cerevisiae chromosome IX cosmid 9877., N = 2, Score = 126, P = 7.1e-09

PIR:A34691 meiosis-specific protein HOP1 - yeast (Saccharomyces cerevisiae), N = 2, Score = 126, P = 7.8e-08

>TREMBL:ATAC2130_3 product: "F1N21.3"; The sequence of BAC F1N21 from Arabidopsis thaliana chromosome 1, complete sequence.
Length = 562

HSPs:

Score = 274 (41.1 bits), Expect = 5.7e-22, P = 5.7e-22
Identities = 84/290 (28%), Positives = 145/290 (50%)

```

Query:   22 TEHQSLVLVKRLLAHSVSCITYLRGIFPECAYGTRYLDDLCVKILREDKNCPGSTQLVKW 81
          TE  SL+L + LL +++ I+Y+RG+FPE + + + L +KI + S +L+ W
Sbjct:   11 TEQDSLLLTRNLLRIAIFNISYIRGLFPEKYFNDKSVPALDMKIKLMPMDAESRLIDW 70

Query:   82 M-LGCYDALQKKYVYT-----NPEDPQTISECYQFKFYTNNGP--LMDFISK--NQSN 130
          M  G YDALQ+KY+ T D I E Y F F Y+++ +M I++ N+ N
Sbjct:   71 MEKGVYDALQRKYLKTLMFISICETVDGPMIEE-YSFYSYSDSDSQDVMNMNINRTGNKK 129

Query:   131 ESSMLST-----DTKKASILLIRKIYILMQNLGPLPNDVCLTMKLFYYDEVTPPDYQPP 184
          ST + + + +R + LM+ L +P++ + MKL YYD+VTPPDY+PP
Sbjct:   130 GGIFNSTADITPNQMRSSACKMVRTLVQLMRTLDKMPDERTIVMKLLYDDVTPPDYEP 189

Query:   185 GFKD--GDCEGVIFEGEPMYLNVGEVSTPFHIFKVKVTT-----ERERENIDSTILS 235
          F+ D ++ P+ + +G V++ + +KV + E + M++ D +
Sbjct:   190 FFRGCTEDEAQYVWTKNPLRMEIGNVNSKHLVLTLVKVKSVLPCEDEDDMQD-DGKSIG 248

Query:   236 PKQIKTPFQKILRDKDVEDEQEHY-----TSDDLDIETKMEEQEKNPASSE 281
          P + Q D ++ QE+ DD D E ++ +PA +E
Sbjct:   249 PDSVHDD-QPSDSDSEISQTQENQFIVAPVEKQDDDDGEVDEDDNTQDPAENE 300

```

Pedant information for DKFZphtes3_15a13, frame 2

Report for DKFZphtes3_15a13.2

```

[LENGTH]      387
[MW]           44417.64
[pI]           5.57
[HOMOL]        TREMBL:ATAC2130_3 product: "F1N21.3"; The sequence of BAC F1N21 from
Arabidopsis thaliana chromosome 1, complete sequence. 9e-23
[FUNCAT]       09.13 biogenesis of chromosome structure [S. cerevisiae, YIL072w] 7e-11
[FUNCAT]       03.19 recombination and dna repair [S. cerevisiae, YIL072w] 7e-11
[FUNCAT]       03.13 meiosis [S. cerevisiae, YIL072w] 7e-11
[FUNCAT]       30.10 nuclear organization [S. cerevisiae, YIL072w] 7e-11
[PIRKW]        nucleus 2e-09
[PIRKW]        zinc finger 2e-09

```

```

[PIRKW]      DNA binding 2e-09
[PROSITE]    MYRISTYL      1
[PROSITE]    CAMP_PHOSPHO_SITE      3
[PROSITE]    CK2_PHOSPHO_SITE      12
[PROSITE]    PKC_PHOSPHO_SITE      7
[PROSITE]    ASN_GLYCOSYLATION      3
[KW]         Alpha_Beta

SEQ  MATAQLQRTPMSALVFPNKISTEHQSLVLVKRLAVSVSCITYLRGIFPECAYGTRYLDD
PRD  cccccccccccccccccchhhhhhhhhhhhhhhhhhhheeeecccccccccchh

SEQ  LCVKILREDKNCPGSTQLVKWMLGCYDALQKKYVYTNPEDPQTISECYQFKFYTNNGPL
PRD  hhhhhhccccccccccccccccchhhhhhhhhhhccccccccchhhheeeeccccce

SEQ  MDFISKNSNESSMLSTDTKKASILLIRKIYILMQNLGPLPNDVCLTMKLFYYDEVTPPD
PRD  eeeeeccccccccceccccchhhhhhhhhhhhhhhhhccccccccceeeeeeecccccc

SEQ  YQPPGFGDGDCEGVIFEGEPMYLVNVEVSTPFHIFKVKVTTTEREREMIDSTILSPKQIK
PRD  cccccccccccccceeeecceeeecccccccccceeeeccccchhhhhccccccccchh

SEQ  TPFQKILRDKDVEDEQEHYTSDDLDIETKMEEQKNPASSELEEPSLVCEEDEIMRSKES
PRD  hhhhhhhhhhhhhhhhhccccchhhhhhhhhhhccccccccccccchhhhhhhhhcc

SEQ  PDLISISHSQEQLVNKTSSELDMSSEKTRSGKVFQNMANGNPVKSSKENRKRSQHESGR
PRD  cccccchhhhhhhhhccccccccccccceeeecccccccccchhhhhhhhhccccce

SEQ  IVLHHFDSSSQESVPKRRKFSEPKHEI
PRD  eeeeecccccccccccccccccccccc

```

Prosites for DKFZphtes3_15a13.2

PS00001	127->131	ASN_GLYCOSYLATION	PDOC00001
PS00001	130->134	ASN_GLYCOSYLATION	PDOC00001
PS00001	315->319	ASN_GLYCOSYLATION	PDOC00001
PS00004	140->144	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	351->355	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	378->382	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	139->142	PKC_PHOSPHO_SITE	PDOC00005
PS00005	167->170	PKC_PHOSPHO_SITE	PDOC00005
PS00005	221->224	PKC_PHOSPHO_SITE	PDOC00005
PS00005	235->238	PKC_PHOSPHO_SITE	PDOC00005
PS00005	329->332	PKC_PHOSPHO_SITE	PDOC00005
PS00005	346->349	PKC_PHOSPHO_SITE	PDOC00005
PS00005	358->361	PKC_PHOSPHO_SITE	PDOC00005
PS00006	96->100	CK2_PHOSPHO_SITE	PDOC00006
PS00006	103->107	CK2_PHOSPHO_SITE	PDOC00006
PS00006	177->181	CK2_PHOSPHO_SITE	PDOC00006
PS00006	221->225	CK2_PHOSPHO_SITE	PDOC00006
PS00006	260->264	CK2_PHOSPHO_SITE	PDOC00006
PS00006	268->272	CK2_PHOSPHO_SITE	PDOC00006
PS00006	280->284	CK2_PHOSPHO_SITE	PDOC00006
PS00006	308->312	CK2_PHOSPHO_SITE	PDOC00006
PS00006	318->322	CK2_PHOSPHO_SITE	PDOC00006
PS00006	346->350	CK2_PHOSPHO_SITE	PDOC00006
PS00006	354->358	CK2_PHOSPHO_SITE	PDOC00006
PS00006	369->373	CK2_PHOSPHO_SITE	PDOC00006
PS00008	84->90	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_15a13.2)

DKFZphtes3_15c24

group: metabolism

DKFZphtes3_15c24 encodes a novel 404 amino acid protein with strong similarity to 2-hydroxyacid dehydrogenases.

The novel protein contains a D-isomer specific 2-hydroxyacid dehydrogenases signature. Proteins with such a signature have similar enzymatic activities: D-lactate dehydrogenase (EC 1.1.1.28), catalyzes the reduction of D-lactate to pyruvate. D-glycerate dehydrogenase (EC 1.1.1.29) catalyzes the reduction of hydroxypyruvate to glycerate. 3-phosphoglycerate dehydrogenase (EC 1.1.1.95), catalyzes the oxidation of D-3-phosphoglycerate to 3-phosphohydroxypyruvate. Therefore the novel protein is a new 2-hydroxyacid dehydrogenase.

The new protein can find application in modulation of 2-hydroxyacid dehydrogenases-dependent pathways and as a new enzyme for biotechnologic production processes.

strong similarity to C.elegans T03F1.1

potential start at Bp 55 matches kozak consensus PyCCatgG

Sequenced by GBF

Locus: unknown

Insert length: 1956 bp

Poly A stretch at pos. 1929, polyadenylation signal at pos. 1903

```
1 CGAAGGCGGC GCGGAAGGCC CGGGCTGGGA GCGTTGGCGG CCGGAGTCCC
51 AGCCATGGCG GAGTCTGTGG AGCGCCTGCA GCAGCGGGTC CAGGAGCTGG
101 AGCGGGGAAC TGCCCAAGGAG AGGAGTCTGC AGGTCCCGAG GAGCGGCGAC
151 GGAGGGGGCG GCCGGGTCCG CATCGAGAAG ATGAGCTCAG AGGTGGTGGA
201 TTCGAATCCC TACAGCCGCT TGATGGCATT GAAACGAATG GGAATTGTAA
251 GCGACTATGA GAAAATCCGT ACCTTTGCCG TAGCAATAGT AGGTGTTGGT
301 GGAGTAGGTA GTGTGACTGC TGAATGCTG ACAAGATGTG GCATTGGTAA
351 GTTGCTACTC TTTGATTATG ACAAGGTGGA ACTAGCCAAT ATGAATAGAC
401 TTTTCTTCCA ACCTCATCAA GCAGGATTAA GTAAAGTTCA AGCAGCAGAA
451 CATACTCTGA GGAACATTAA TCCTGATGTT CTTTTTGAAG TACACAATA
501 TAATATAACC ACAGTGGAAC ACTTCAACA TTTCATGGAT AGAATAAGTA
551 ATGGTGGGTT AGAAGAAGGA AAACCTGTTG ATCTAGTTCT TAGCTGTGTG
601 GACAAATTTG AAGCTCGAAT GACAATAAAT ACAGCTTGTA ATGAACTTGG
651 ACAAACATGG ATGGAATCTG GGGTCAGTGA AAATGCAGTT TCAGGGCATA
701 TACAGCTTAT AATTCCTGGA GAATCTGCTT GTTTTGGCTG TGCTCCACCA
751 CTGTAGTTG CTGCAATAT TGATGAAAAG ACTCTGAAAC GAGAAGGTGT
801 TTGTGCAGCC AGTCTTCCCTA CCACTATGGG TGTGGTTGCT GGGATCTTAG
851 TACAAACAGT GTTAAAGTTT CTGTTAAATT TTGGTACTGT TAGTTTTTAC
901 CTGGATACA ATGCAATGCA GGATTTTTTT CCTACTATGT CCATGAAGCC
951 AAATCCTCAG TGTGATGACA GAAATTGCAG GAAGCAGCAG GAGGAATATA
1001 AGAAAAAGGT AGCAGCACTG CCTAAACAAG AGGTTATACA AGAAGAGGAA
1051 GAGATAATCC ATGAAGATAA TGAATGGGTT ATTGAGCTGG TATCTGAGGT
1101 TTCAGAAAGG GAACTGAAAA ATTTTTCAGG TCCAGTTCCA GACTTACCTG
1151 AAGGAATTAC AGTGGCATA CAAATTCCTA AAAAGCAAGA AGATTCTGTC
1201 ACTGAGTTAA CAGTGAAGA TTCTGGTGAA AGCTTGGAAG ACCTCATGGC
1251 CAAATGAAG AATATGTAGA TAATGGACTG GGATATATTG TATTTCTCAT
1301 GTTAAAGCCT CTCCCTTGA AATTAAAAAA AAATTTTAAC TGATAAACT
1351 TAGGGCAACA TTAATTAATG TATATTCTTA CCTGAATTGT TATACTTTTT
1401 GAAATCCTG TGAATTCCTT GTTCTCCCC GCTCCAACGA AATCATTAAC
1451 TCTCTAAAA TGTGTTTCAT TCTAGTAAGA AAACCTCAAA GGATATTGTA
1501 GGATATAAAT CTTACTTGAA AACATAGCTG TTGAAATGTT TTGGCCTTTT
1551 GGAGTGGGGG AAGGACAAAT CTGATCCTGT AATCTTTTTC TTTCCAGTAA
1601 TCCCTTGTGT CTGTTGCATG AGGACATGGA CAATAAAGTA GTATATGATC
1651 CTCAGATACA GGGAGAAGGA CAAGGCATAC AGCTTATTGA TTAGAGCTGG
1701 CAAGCATCTG CTCATTATGT TTGGAATTGC TTTCTATAAG AAAATTGCCC
1751 ACTACTACTA ACTTGATCAA CAATGAATTC AAAATAGTTA ACCTATGAAA
1801 TAACATCCTC TCAAAATGTT GCTGATGAAG TACAAGTTGA AATGTAGTTA
1851 TTGGAAAAGT CTGTAACCTG TGGATCATAT ATATTCAAAG TGAGACAAAG
1901 GCAAAATAAA AGCAGCTATT TTCATGAATA GACAAAAAAA AAAAAAAG
1951 AAAAAAG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 55 bp to 1266 bp; peptide length: 404

Category: similarity to unknown protein

Classification: Metabolism

Prosite motifs: D_2_HYDROXYACID_DH_1 (76-105)

```

1 MAESVERLQQ RVQELERELA QERSLQVPRS GDGGGGRVRI EKMSSEVVDS
51 NPYSRLMALK RMGIVSDYEK IRTFAVAIVG VGGVGSVTAE MLTRCGIGKL
101 LLFDYDKVEL ANMNLFFQF HQAGLSKVQA AEHTLRNINP DVLFEVHNYN
151 ITTVENFQHF MDRISNGGLE EGKPVDLVLS CVDNFEARMT INTACNELGQ
201 TWMESGVSEN AVSGHIQLII PGESACFACA PPLVVAANID EKTLLKREGVC
251 AASLPTTMGV VAGILVQNVL KFLNFGTVS FYLGYNAMQD FFPTMSMKPN
301 PQCDNRNCRK QQEYKKKVA ALPKQEVQIE EEEIHEDNE WGIELVSEVS
351 EEELKNFSGP VPDLPegITV AYTIKKQED SVTELTVEDS GESLEDLMK
401 MKNM

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_15c24, frame 1

TREMBL:CEUT03F1_11 gene: "T03F1.1"; Caenorhabditis elegans cosmid
T03F1., N = 1, Score = 1204, P = 1.9e-122

TREMBL:ATAC98_3 gene: "YUP8H12.3"; Arabidopsis thaliana chromosome 1
YAC yUP8H12 complete sequence., N = 1, Score = 733, P = 1.5e-72

PIR:A69319 thiamine biosynthesis protein (thiF) homolog - Archaeoglobus
fulgidus, N = 1, Score = 218, P = 1.8e-17

TREMBL:AF022796_4 gene: "moeB"; product: "MoeB"; Staphylococcus
carneus molybdenum cofactor biosynthetic gene cluster, complete
sequence., N = 1, Score = 220, P = 3.7e-16

>TREMBL:CEUT03F1_11 gene: "T03F1.1"; Caenorhabditis elegans cosmid T03F1.
Length = 419

HSPs:

Score = 1204 (180.6 bits), Expect = 1.9e-122, P = 1.9e-122
Identities = 241/367 (65%), Positives = 293/367 (79%)

```

Query:   37 RVRIEKMSSEVVDSNPYSRLMALKRMGIVSDYEKIRTFVAIVGVGGVGSVTAEMLTRCG 96
          R +IEK+S+EVVDSNPYSRLMAL+RMGIV++YE+IR  VA+VGVGGSV AEMLTRCG
Sbjct:   48 RQKIEKLSAEVVDSNPYSRLMALQRMGIVNEYERIREKTAVVGVGGVGSVVAEMLTRCG 107

Query:   97 IGKLLLFDDYDKVELANMNLFFQPHQAGLSKVQA AEHTLRNINPDVLFEVHNYNITTVEN 156
          IGKL+LFDDYDKVE+ANMNLRF+QP+QAGLSKV+AA  TL ++NPDV  EVHN+NITT++N
Sbjct:  108 IGKLILFDYDKVEIANMNLRFYQPNQAGLSKVEAARDTLIHVNPDVQIEVHNFNITMDN 167

Query:  157 FQHFMDRISNGGLEEGKPVDLVLSVDNFEARMTINTACNELGQTMWESGVSENAVSGHI 216
          F F++RI  G L +GK +DLVLSVDNFEARM +N ACNE  Q WMESGVSENAVSGHI
Sbjct:  168 FDTFVNRIKRGSLTDGK-IDLVLSVDNFEARMVNMACEENQIWMESGVSENAVSGHI 226

Query:  217 QLIIPGESACFACAPPLVVAANIDEKTLKREGVCAASLPTTMGVVAGILVQNVLKFLNLF 276
          Q I PG++ACFAC PPLVVA+ IDE+TLKR+GVCAASLPTM VVAG LV N LK+LLNF
Sbjct:  227 QYIEPGKTACFACVPPLVVASGIDERTLKRQGVCAASLPTTMVAVAGFLVMNTLKYLLNF 286

Query:  277 GTVSFYLGYNAMQDFFPTMSMKPNPQCDNRNCRKQEEYKKKVAALPKQ-EV-IQEEEEI 334
          G VS Y+GYNA+ DFFP  S+KPNP CDD +C ++Q+EY++KVA  P  EV + EEE +
Sbjct:  287 GEVSQYVGNALSDFFPRDSIKPNPYCDDSHCLQRQKEYEEKVANQPVLDLEVEVP EETV 346

Query:  335 IHEDNEWGIELVSEVSEELKNFSGPVPDLPEGITVAYTIKKQEDSVTELTVEDSGESL 394
          +HEDNEWGIELV+E SE  + S  + G+ AY P K+ D+ TEL+  + +
Sbjct:  347 VHEDNEWGIELVNE-SEPSAEQSSSL--NAGTGLKFAYE-PIKR-DAQTELSPAQA--AT 399

Query:  395 EDLMAMKN 403

```

Report for DKFZphtes3 15c24.1

```

SEQ      MAESVERLQQRVQELERELAQERSLQVPRSGDGGGVRIEKMSSEVVDNSNPYSRLMAK
SEG      .
PRD      ccchhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccceeeccccccccchhhhhhhc
MEM      .....

SEQ      RMGIVSDYEKIRTFVAIVGVGGVGSVTAEMLTRCGIGKLLFDYDKVELANMNRLLFFQP
SEG      .....xxxxxxx
PRD      cccccchhhhhhheeeccccchhhhhhhhhhhhhccccceeeccccchhhhhhhhhhhc
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMMMM

SEQ      HQAGLSKVQAAEHLTRNINPDLFEVHNYNITTVENFQHFMDRISNGGLEEGKPDVLVLS
SEG      .
PRD      cccccchhhhhhhhhhhhhccccceeeccccccccchhhhhhhhhhhhhccccccccceeee
MEM      .....

SEQ      CVDNFEARMTINTACNELGQTMESGVSENAVSGHIQLIIPGESACFACAPPLVVAANID
SEG      .
PRD      cccchhhhhhhhhhhhhhhccccccccccccccccceeecccccccccecccccccccc
MEM      .....

SEQ      EKTCLKREGVCAASLPTTMGVVAGILVQNVLKFLNFGTVSFYLGYNAMQDFFPTMSMKPN
SEG      .
PRD      cccccccccccccccccchhhhhhhhhhhhhhhhhhhhhccccceeecccccccccccccccccc
MEM      .....

SEQ      PQCDNRNCRKQEEYKKKVAALPKQEVIEEEEEIIHEDNEWGIELVSEVSEELKNFSGP
SEG      .....xxxxxxxxxxxxxxxx.....xxxxxxxxxxx
PRD      cccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccceeeehhhhhhhhhcccc
MEM      .....

SEQ      VPDLPPEGITVAYTIPKKQEDSVTELTVEDSGESLEDLMAKMKNM
SEG      .
PRD      cccccceeeeeeohhhhhhhhheeeccccchhhhhhhhhhhcc
MEM      .....

```

(No Pfam data available for DKFZphtes3 15c24.1)

DKFZphtes3_15c6

group: transmembrane protein

DKFZphtes3_15c6 encodes a novel 118 amino acid protein without similarity to known proteins.

The novel protein contains 1 transmembrane region.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes and as a new marker for testicular cells.

unknown

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1283 bp

Poly A stretch at pos. 1264, no polyadenylation signal found

```
1 GAGACACTGA GCCCGAGAC AGTGAGTGGT GGCCTCACTG CTCTGCCCGG
51 CACCCGTGCA CCTCCACTTT GCCTTGTGG AAGTGACCCA GCCCCTCCC
101 CTTCCATTCT CCCACCTGTT CCCAGGACT CACCCAGCC CCGCCTGCC
151 CCTGAGGAAG AAGAGGCACT CACCACTGAG GACTTTGAGT TGCTGGATCA
201 GGGGGAGCTG GAGCAGCTGA ATGCAGAGCT GGGCTTGAG CCAGAGACAC
251 CGCCAAAACC CCCTGATGCT CCACCCCTGG GGGCCGACAT CCATTCTCTG
301 GTACAGTCAG ACCAAGAAGC TCAGGCCGTG GCAGAGCCAT GAGCCAGCCG
351 TTGAGGAAGG AGCTGCAGGC ACAGTAGGGC TTCCTGGCTA GGAGTGTTCG
401 TGTTTCTCTC TTTGCCTACC ACTCTGGGGT GGGGCACTGT GTGGGAAGC
451 TGGCTGTGCG ATGGTAGCTA TTCCACCCTC TGCCTGCCTG CCTGCCTGCT
501 GTCCTGGGCA TGGTGCACTA CCTGTGCCTA GGATTGGTTT TAAATTTGTA
551 AATAATTTTC CATTTGGGTT AGTGGATGTG AACAGGGCTA GGGAAAGTCT
601 TCCACAGGCC TGGCCTTGCC TCCCTGCCTC ATCTCTATTC TCATTCCACT
651 ATGCCCAAG CCCTGGTGGT CTGGCCCTTT CTTTTCCTC CTATCCTCAG
701 GGACCTGTGC TGCTCTGCCC TCATGTCCCA CTTGGTTGTT TAGTTGAGGC
751 ACTTTATAAT TTTTCTCTTG TCTGTGTTC CTTCTGCTT TATTTCCCTG
801 CTGTGTCTCG TCCTTAGCAG CTCAACCCCA TCCTTTGCCA GCTCCTCCTA
851 TCCCGTGGGC ACTGGCCAAG CTTAGGGAG GTCCTGGTC TGGGAAGTAA
901 AGAGTAAACC TGGGGCAGTG GGTGAGGCCA GTAGTTACAC TCTTAGGTCA
951 CTGTAGTCTG TGTAACCTTC ACTGCATCCT TGCCCCATTG AGCCCGGCCT
1001 TTCATGATGC AGGAGAGCAG GGATCCCGCA GTACATGGCG CCAGCACTGG
1051 AGTTGGTGAG CATGTGCTCT CTCTTGAGAT TAGGAGCTTC CTTACTGCTC
1101 CTCTGGGTGA TCCAAGTGTA GTGGGACCCC CTACTAGGGT CAGGAAGTGG
1151 AACTAACAT CTGTGCAGGT GTTGACTTGA AAAATAAAGT GTTGATTGGC
1201 TAAAAAATAA AAAAAAATAA AAAAAAATAA AAGGGCGGCC GCTCTAGAGG
1251 ATCCAAGCTT ACGTAAAAAA AAAAAAATAA AAG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 461 bp to 814 bp; peptide length: 118
Category: putative protein

```
1 MVAIPPSACL PACCPGHGAV PVPRIGFKFV NNFPFGLVDV NRAREVLPTA
51 CACLPAASSLF SFHYAPSPGG LALSFSSYPQ GPVLLCPHVP LGCLVEALYN
101 FSLVLCSFLL YFPAVSCP
```


BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_15c6, frame 2

PIR:S54250 ribosomal protein L2 - Arabidopsis thaliana, N = 1, Score = 76, P = 0.33

>PIR:S54250 ribosomal protein L2 - Arabidopsis thaliana
Length = 258

HSPs:

Score = 76 (11.4 bits), Expect = 4.0e-01, P = 3.3e-01
Identities = 30/91 (32%), Positives = 44/91 (48%)

```
Query:   15 PGHGAVPVPRIGFKFVNNFPFGLVDVNRAREVLPTACACLPASSLFSFHYAPSPGGLALS 74
          PG GA P+ R+ F+   PF   + +E+  A C P SSL+  A   G L
Sbjct:   52 PGRGA-PLARVTFRH----PFRF---KKQKELFVAAEVCTPVSSLYCGKKATLVVGNVLP 103

Query:   75 FSSYPQGPVLLCP---HV-PLGCLVEALYNFSLVL 105
          S P+G V+ C   HV   G L A ++++V+
Sbjct:  104 LRSIPEGAVV-CNVEHHVGDGVLARASGDYAIVI 137
```

Pedant information for DKFZphtes3_15c6, frame 2

Report for DKFZphtes3_15c6.2

```
[LENGTH]      118
[MW]           12413.79
[pI]           7.53
[PROSITE]      LEUCINE_ZIPPER 1
[PROSITE]      MYRISTYL 1
[PROSITE]      ASN_GLYCOSYLATION 1
[KW]           TRANSMEMBRANE 1
```

```
SEQ      MVAIPPSACLPAACCPGHGAVPVPRIGFKFVNNFPFGLVDVNRAREVLPTACACLPASSLF
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM      .....

SEQ      SFHYAPSPGGLALSFSSYPQGPVLLCPHVPLGCLVEALYNFSLVLCSEFLLYFPAVSCP
PRD      eeeeecccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM      .....MMMMMMMMMMMMMMMMMM.
```

Prosites for DKFZphtes3_15c6.2

```
PS00001    100->104  ASN_GLYCOSYLATION    PDOC00001
PS00008     70->76   MYRISTYL          PDOC00008
PS00029     84->106   LEUCINE_ZIPPER    PDOC00029
```

(No Pfam data available for DKFZphtes3_15c6.2)

DKFZphtes3_15g14

group: testes derived

DKFZphtes3_15g14 encodes a novel 701 amino acid protein with weak similarity to *S. cerevisiae* hypothetical protein YOR243c.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to YOR243c

complete cDNA, complete cds, potential start codon at Bp 35, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 3495 bp

Poly A stretch at pos. 3462, no polyadenylation signal found

```

1 GCCTTCCACT GAACCGAGGC ACTGTTATAG AAGAATGGAA GAAGATACAG
51 ATTATAGAAT CAGGTTTAGT TCTTTGTGTT TCTTTAATGA TCACGTTGGA
101 TTTTCATGGCA CTATAAAAAG CTCACCAAGT GACTTTATTG TTATTGAAAT
151 TGATGAACAG GGACAGTTAG TTAATAAGAC CATCGATGAG CCTATTTTCA
201 AGATTAGTGA AATACAACCT GAGCCAAATA ATTTCCCAA AAAACCAAAA
251 CTAGATCTTC AAAATCTGTC CTTAGAAGAT GGAAGAAACC AAGAAGTTCA
301 TACTTTGATT AAGTACACTG ATGGTGACCA AAATCATCAG TCTGGTTCAG
351 AAAAGGAAGA TACTATCGTT GATGGAACCT CCAATGTGA AGAAAAAGCT
401 GATGTTTTAA GCTCCTTTT GGATGAAAAA ACTCATGAGT TACTGAATAA
451 TTTTGCTGTG GATGTAAGAG AGAAGTGGCT TTCTAAAACA GAGCTAATTG
501 GACTACCTCC TGAATTCTCA ATAGGCAGAA TCCTTGACAA AAACCGAGAG
551 GCTAGTTTAC ACAGTGCCAT TAGGCAGAAA TTTCCATTTT TAGTAAGTGT
601 AGGAAAAAAC AGTGAAATTG TTGTAAAACC AAATCTTGAA TATAAAGAAC
651 TTTGTCATTT GGTATCTGAA GAGGAAGCAT TTGACTTTT TAAATATTTG
701 GATGCAAGA AAGAAAATTC CAAATTTACC TTTAAACCTG ATACAAACAA
751 AGACCACAGA AAGCTGTCC ACCATTTTGT CAACAAAAG TTTGGAACCC
801 TTTGGAAGAC CAAATCTTT TCTAAAATGA ATTGCAGTGC TGGTAATCCG
851 AATGTGGTGG TAACAGTAAG ATTTCGGGAA AAGCACACA AACGTGGGAA
901 AAGGCCTCTT TCTGAATGCC AAGAAGGAAA AGTTATATAT ACAGCTTTTA
951 CCCTAGGAAA GGAACCTG GAAATGTTT AAGCGATTGG TTTTTTAGCT
1001 ATCAAACTTG GTGTTATTCC TTCGGATTT AGTTATGCG GCCTTAAAGA
1051 CAAGAAAGCC ATCACCTATC AAGCAATGGT TGTTAGAAA GTGACTCCAG
1101 AGAGGTTGAA AAATATTGAA AAAGAAATTG AAAAGAAAAG AATGAATGTC
1151 TTTAATATT CCGTCTGTAG TGATTCCCTG AGACTTGGTC AGCTCAAAGG
1201 AAATCACTTT GATATTGTCA TTGAAAATTT AAAAAACAA ATAAATGATT
1251 CTGCAAACTT GAGGGAGAGA ATTATGGAAG CAATAGAAA TGTTAAGAAA
1301 AAAGGCTTTG TGAATTACTA TGGACCACAG AGATTTGGGA AGGGAAGGAA
1351 AGTTACACAC GACCAATTG GACTAGCTTT GCTGAAGAAT GAAATGATGA
1401 AAGCCATAAA ATTGTTTCTT ACACCAGAAG ACTTGATGA TCCTGTAAT
1451 AGAGCAAGA AGTATTTTCT TCAAACTGAG GATGCTAAAG GCACACTTTC
1501 ATTGATGCCT GAATTCAAAG TCGGTGAGAG AGCATTGTTG GAGGCATTGC
1551 ACCGCTTTGG CATGACCGAG GAAGGTTGTA TCCAGGCATG GTTCTCTTAA
1601 CCCCATTCCA TGCGCATATT CTATGTTTAC GCATATACCA GCAAAATTTG
1651 GAATGAGGCA GTATCTTACA GACTTGAAAC CTATGGAGCA AGAGTAGTGC
1701 AGGGTGATTT GGTCTGTTTG GATGAAGACA TTGATGACGA GAATTTCCCA
1751 AATAGTAAAA TTCACCTGGT AACTGAAGAG GAGGGATCAG CTAATATGTA
1801 TGCAATACAT CAGGTGGTTC TTCCAGTACT TGGATACAAT ATTCAGTACC
1851 CGAAGAACAA AGTAGGCAG TGGTACCATG ACATACTTAG CAGAGATGGA
1901 CTACAGACAT GTAGGTTTAA AGTACCTACT CTGAACTGA ATATACCAGG
1951 TTGCTATAGA CAGATTTTGA AACATCCCTG TAATCTCTCA TACCAACTAA
2001 TGAAGATCA TGACATTGAT GTCAAAACGA AAGGTTCCCA CATTGATGAA
2051 ACAGCTTTGT CTCTTTTGAT CTCTTTTGAT CTTGATGCTT CATGCTATGC
2101 TACCGTTTGT CTGAAGGAAA TAATGAAGCA TGACGTTTAA AACTGATACC
2151 CTTGGTATAA CCATATATAT GTCACCCTTT CTTGTTTTTG AAATTATTGA
2201 TCAGACAAT ATACAAGGGA AATGCCATAC CTCTGTTTGT GATAGATACC
2251 CCAGAGTAGT TATTACCTCT TTGTGAGATA AGTAATCTTT GATGAAGATT
2301 GAAATACAAT TTCTCATCCA ATTTTATAT CTTGGCATAC GCTGACCCTC
2351 TTGACCATTT GTAATTTTT CATATTATCT AAAACAGGTG TTAGAGTCAG
2401 ACAGATTCA TCTTAGATT TAGCTCTGAC ACTTACTAGT GATTTGAGT
2451 ATGTTGTTGA TTTTTTGTG TGTGGTTACT GATAGAATCA AGACAATTAC
2501 AACTTCATAA ATGACAAATA ATAGGATTAT CTCACATTT TCTGTTGCTG
2551 GAGGAACAAA ACATTGTGCC CATTGAAAA TTTTAATTT TGTGGTTTAA
2601 ACTATCCAC ATTATAAATC ATCCTTACC ATTTATATC AGTTAAATAT
2651 GGGTGTGTTG GGGAGGAATG ACTGGCATGT AGACATGTAT TGATTTAGGA
2701 AGATCTGAGC ATTTCTTCA TTGTTGTTAA GATATAATGA TGAATTTAA

```

```

2751 AAAGCAGTAT GGAGCATTAT ATATCAGTAA TGTGATATAT ATACTTAAGC
2801 CAGTTTAAACC ATTTTGGGAA ATGTTAGCAT TAGGAAATAA AATCCAAAAG
2851 AAGGAAGAGA AGCTATATGC AATGCAAAAT TTGCTTATTG CAATATTTTC
2901 ATATACAGAC ACTAAAAACA GTTTTCAAAG TCCAGCATTG CGTAACTAAA
2951 GTAAGTAAAA TGATGTGTAT CAACTTGATG GTAAAAATATG TAGTTATTTA
3001 AAAAAGCAAT GAACAATTTA GTTTCATGAG AAAATGTTGC CCCCTAAAAG
3051 TAGAACACAT ATGTTACAAC TGCAATAATA CTCTGAATTC ATCTTTCACA
3101 AATAAGAGAC ATGTTAGCAT AGTGATTAAA AGCACAGATA TTGGAGACAA
3151 ACTAAGCCAG TTTGAACCCCT GGCACCTGCCA CGTATAGCAC TGCAGCCTTG
3201 GGAAAGTTAT TTAAACTCAT GGGCTTCAGT TTCAACATCT GTAAATGCGG
3251 CATGTTAACA TTGCCTACCT CATAGGATTA CTGTGAGAAT TTTCTAAGTT
3301 AATATATGTA AAGCAACTTT AAAAAGTGCC TGGCACTTAG TTATTGTTAA
3351 GTAAGTGTCT GCAGATGCAA GTTTGGAAGA GAAAAGCAA TAAATGAAAA
3401 TCCCTTCCTG TTAAGATGAA AAAAAAAAAA AAAAAAAAAA AAAAAAGGGG
3451 CGGCCGCTCA AGATGAAAAA AAAAAAAAAA AAAAAAAAAA AAAGG

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 35 bp to 2137 bp; peptide length: 701
 Category: similarity to unknown protein

```

1 MEEDTDYRIR FSSLCFFNDH VGFHGTIKSS PSDFIVIEID EQGQLVNKTI
51 DEPIFKISEI QLEPNNFPPK PKLDLQNLISL EDGRNQEVHT LIKYTDGDQN
101 HQSGSEKEDT IVDGTSKCEE KADVLSSFLD EKTHELLNMF ACQVREKWS
151 KTELIGLPPE FSGRILDKN QRASLHSAIR QKFPPFLVTVG KNSEIVVKPN
201 LEYKELCHLV SEEEAFDFFK YLDAKKENSK FTFKPDNTKD HRAKAVHHFVN
251 KKEGNNLVETK SFSKMNCASG NPNVVTVRF REKAHKGKR PLSECEGKV
301 IYTAFTLRKE NLEMFEAIGF LAIKLGVIPs DFSYAGLKDK KAITYQAMVV
351 RKVTPERLKN IEKEIEKKRM NVFNIRSVDD SLRLGQLKGN HFDIVIRNLK
401 KQINDSANLR ERIMEAIENV KKGGFVNYG PQRFKGGRKV HTDQIGLALL
451 KNEEMKAIKL FLTPEDLDDP VNRKKYFLQ TEDAKGTLISL MPEFKVRERA
501 LLEALHRFGM TEEGCIQAWF SLPHSMRIFY VHAYTSKIWN EAVSYRLETY
551 GARVVQGDV CLDEDIDEN FPNSKIHLVT EEEGSANMYA IHQVVLPLVG
601 YNIQYPKNKV GQWYHDILSR DGLQTCRFKV PTLKLNIPGC YRQILKHPCN
651 LSYQLMEDHD IDVKTGSHI DETALSLLIS FDLASCYAT VCLKEIMKHD
701 V

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_15g14, frame 2

TREMBL:SPBC1A45P_10 gene: "SPBC1A4.09"; product: "hypothetical protein"; S.pombe chromosome II cosmid clA4 left hand region 1-26184 bp
 Originates from chimeric cosmid., N = 3, Score = 511, P = 2.9e-57

PIR:S67136 hypothetical protein YOR243c - yeast (Saccharomyces cerevisiae), N = 2, Score = 516, P = 7.3e-54

SWISSPROT:YQ4B_CAEEL HYPOTHETICAL 64.6 KD PROTEIN B0024.11 IN CHROMOSOME V., N = 2, Score = 386, P = 2.1e-34

>PIR:S67136 hypothetical protein YOR243c - yeast (Saccharomyces cerevisiae)
 Length = 676

HSPs:

Score = 516 (77.4 bits), Expect = 7.3e-54, Sum P(2) = 7.3e-54
 Identities = 151/498 (30%), Positives = 245/498 (49%)

Query: 191 KNSEIVVKPNLEYKELCHLVSEEEAFDFFK-YLDAKKENSKFTFKPDNTKDHRAKAVHHFV 249
 + E V P L +L + EE+ Y A K + F+ +K R +H +

Sbjct: 109 RRQEFNVDPPELR-NQLVEIFGEEDVLKIESVYRTANKMETAKNFE---DKSVRTKIHQLL 164

Query: 250 NKKFGNVLVETKFSKMNCSAGNPVNVTVRFREKAHK-RGKRPLSECQEG-KVIYTAFTL 307
+ F N +E+ + N +EK ++ R + G + FTL

Sbjct: 165 REAFKNELESVTTDTNTFKIARSNRNSRTNKQEKINQTRDANGVENWGYGSPKDFIHF 224

Query: 308 RKENLEMFEAIGFLAIGLVIPSD-FSYAGLKDKKAITQAMVVRKVTPERLKNIEKEIE 366
KEN + EA+ + KL +PS YAG KD++A+T Q + + K+ +RL + + +

Sbjct: 225 HKENKDTMEAVNVIT-KLLRVPSRVIRYAGTKDRAVTCQQRVSISKIGLDRLNALNRTL- 282

Query: 367 KKRNVNFNIRSVDDSLRLGQLKGNHFDIVIRNLKKQINDSANLRERIMEAIENVKKKGFV 426
K M + N D SL LG LKGN F +VIR++ N +L E + + + + GF+

Sbjct: 283 -KGMIIIGNYFSDASLNLGDLKGNFVVVIRDVTTG-NSEVSLEEIVSNGCKSLSENGFI 340

Query: 427 NYYGPRQFGRKRVHTDQIGLALLKNEMKAIKFLTPEDLDDPVNR-AKKYFLQTEDAK 485
NY+G QRFG + T IG LL + KA +L L+ +D P ++ A+K + +T+DA

Sbjct: 341 NYFGMQRFQTF-SISTHTIGRELLSNWKKAAELILSDQDNVLPKSKARKIWAETKDA 399

Query: 486 GTLSLMPEFKVRERALLEALHRFGMTEEGCIQ---AWFS----LPHSMRIFYVHAYTSKI 539
L MP + E ALL +L E+G A+++ +P ++R YVHAY S +W

Sbjct: 400 LALKQMPRQCLAENALLYSLSNQRKEEDGTYSENAYTAIMKIPRNLRTMYVHAYQSYVW 459

Query: 540 NEAVSYRLETYGARVVQDGLVC-----LDEDIDDENFPNS-----KIHVLTEEEGS 585
N S R+E +G ++V GDLV L IDDE+F + VT+E+

Sbjct: 460 NSIASKRIELHGLKLVGDLVIDTSEKSPDISGIDDEDFEDVREAQFIRAKAVTQEDID 519

Query: 586 ANMYAIHQVVLVPLGYNIQYPKNK-VGQWYHDILSRDGLQTCRFKVPTLKLNIPEGYRQI 644
+ Y + VVLP G+++ YP N+ + Q Y DIL D + + ++ G YR +

Sbjct: 520 SVKYTMEDVVLPSPGFDVLYPSNEELKQLYVDILKADNMDFNMRRKVRDFSLAGSYRTV 579

Query: 645 LKHPCNLSYQLMEDHDIDVKTGSHID 671
++ P +L Y+++ D + + +D .

Sbjct: 580 IQKPKSLEYRIIHYDDPSQQLVNTDLD 606

Score = 86 (12.9 bits), Expect = 3.2e-01, Sum P(2) = 2.8e-01
Identities = 40/160 (25%), Positives = 77/160 (48%)

Query: 22 GFHGTIKSSPSDFIVIEIDEQGLVNKTIDEPIFKISEIQLEPNFPKPKLQNLNLSLE 81
GF G IK +DF+V EID++G++++ T D+ FK+ + +P K +++ + S E

Sbjct: 55 GFRGQTKQRYTDFLVNEIDQEGKVIHLT-DKG-FKMPK---KPQR--SKEEVNAEKES-E 106

Query: 82 DGRNQEVHTLIKTYTDGDQNHQSGS--EKEDTI-VDGTSCKEKADVLSSFLDEKTHELLN 138
R QE + D + +Q +ED + ++ + K + +F D+ ++

Sbjct: 107 AARRQEFNV-----DPELRNQLVEIFGEEDVLKIESVYRTANKMETAKNFEDKSVRTKI 161

Query: 139 NFACDVREKWLSTELIGLPPE-FSIGRILDKNQRASLHSAIRQ 181
+RE + ++ E + F I R ++N R + I Q

Sbjct: 162 QL---LREAFKNELESVTTDTNTFKIARS-NRNSRTNKQEKINQ 201

Score = 58 (8.7 bits), Expect = 7.3e-54, Sum P(2) = 7.3e-54
Identities = 10/23 (43%), Positives = 17/23 (73%)

Query: 676 SLLISFDLDASCYATVCLKEIMK 698
++++ F L S YAT+ L+E+MK

Sbjct: 638 AVVLKQFQGLTSAYATMALRELMK 660

Pendant information for DKFZphtes3_15g14, frame 2

Report for DKFZphtes3_15g14.2

[LENGTH] 701
[MW] 80700.96
[pI] 7.31
[HOMOL] PIR:S67136 hypothetical protein YOR243c - yeast (*Saccharomyces cerevisiae*) 2e-51
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YOR243c] 8e-53
[BLOCKS] BL01268C
[BLOCKS] BL01268B
[BLOCKS] BL01268A
[SUPFAM] hypothetical protein HI0701 3e-06
[PROSITE] MYRISTYL 7
[PROSITE] AMIDATION 2
[PROSITE] CAMP_PHOSPHO_SITE 1
[PROSITE] CK2_PHOSPHO_SITE 16
[PROSITE] TYR_PHOSPHO_SITE 1
[PROSITE] PKC_PHOSPHO_SITE 13
[PROSITE] ASN_GLYCOSYLATION 5
[KW] Alpha_Beta

```

SEQ MEEDTDYRIRFSSLCFFNDHVGFGHTIKSSPSDFIVIEIDEQGQLVNKTIDEPIFKISEI
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ QLEPNFPPKPKLDLQNLSDGRNQEVHTLIKYTGDQNHQSGSEKEDTIVDGTSKCEE
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccch

SEQ KADVLSSFLDEKTHELLNNFACDVREKWLKSTELIGLPPEFSIGRILDKNQRASLHSAIR
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ QKFPPFLVTVGKNSEIVVKPNLEYKELCHLVSEEEAFDFKYLDACKENSKFTEKPDNTKD
PRD hhccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccch

SEQ HRKAVHHFVNKKFGNLVETKSFSSKMNCSAGNPNVVTVRFREKAHKGKRLPSECQEGKV
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcc

SEQ IYTAFTLRKENLEMEFAIGFLAIKLGVIPSDFSYAGLKDKKAITTYQAMVVRKVTPERLKN
PRD eeeeeccccccccccccccccccccccccccccccccccccccccccccccccccccccccch

SEQ IEKEIEKKRMNVFNIRSVDDSLRLGQLKGNHFDIVIRNLKKQINDSANLRERIMEAIENV
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ KKKGFVNYYPQRFQKGRKVHTDQIGLALLKNEMMKAIKFLTPEDLDDPVNRKAKYFLQ
PRD hhccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccch

SEQ TEDAKGTLSLMPEFKVRERALLEALHRFGMTEEGCIQAWFSLPHSMRIFYVHAYTSKIWN
PRD hccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccch

SEQ EAVSYRLETYGARVVQGDVLVCLDEDIDDENFPNSKIHLVTEEGSANMYAIHQVVLPLVG
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcc

SEQ YNIQYPKNKVGQWYHDILSRDGLQTCRFKVPKLNIPGCRYQILKHPCNLSYQLMEDHD
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccch

SEQ IDVKTGSHIDETALLISFDLDASCYATVCLKEIMKHDV
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

Prosites for DKFzphes3_15g14.2

PS00001	47->51	ASN_GLYCOSYLATION	PDOC00001
PS00001	77->81	ASN_GLYCOSYLATION	PDOC00001
PS00001	266->270	ASN_GLYCOSYLATION	PDOC00001
PS00001	404->408	ASN_GLYCOSYLATION	PDOC00001
PS00001	650->654	ASN_GLYCOSYLATION	PDOC00001
PS00004	351->355	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	26->29	PKC_PHOSPHO_SITE	PDOC00005
PS00005	105->108	PKC_PHOSPHO_SITE	PDOC00005
PS00005	115->118	PKC_PHOSPHO_SITE	PDOC00005
PS00005	232->235	PKC_PHOSPHO_SITE	PDOC00005
PS00005	237->240	PKC_PHOSPHO_SITE	PDOC00005
PS00005	277->280	PKC_PHOSPHO_SITE	PDOC00005
PS00005	306->309	PKC_PHOSPHO_SITE	PDOC00005
PS00005	381->384	PKC_PHOSPHO_SITE	PDOC00005
PS00005	525->528	PKC_PHOSPHO_SITE	PDOC00005
PS00005	535->538	PKC_PHOSPHO_SITE	PDOC00005
PS00005	544->547	PKC_PHOSPHO_SITE	PDOC00005
PS00005	625->628	PKC_PHOSPHO_SITE	PDOC00005
PS00005	632->635	PKC_PHOSPHO_SITE	PDOC00005
PS00006	30->34	CK2_PHOSPHO_SITE	PDOC00006
PS00006	49->53	CK2_PHOSPHO_SITE	PDOC00006
PS00006	79->83	CK2_PHOSPHO_SITE	PDOC00006
PS00006	95->99	CK2_PHOSPHO_SITE	PDOC00006
PS00006	103->107	CK2_PHOSPHO_SITE	PDOC00006
PS00006	105->109	CK2_PHOSPHO_SITE	PDOC00006
PS00006	110->114	CK2_PHOSPHO_SITE	PDOC00006
PS00006	116->120	CK2_PHOSPHO_SITE	PDOC00006
PS00006	127->131	CK2_PHOSPHO_SITE	PDOC00006
PS00006	150->154	CK2_PHOSPHO_SITE	PDOC00006
PS00006	211->215	CK2_PHOSPHO_SITE	PDOC00006
PS00006	237->241	CK2_PHOSPHO_SITE	PDOC00006
PS00006	377->381	CK2_PHOSPHO_SITE	PDOC00006
PS00006	463->467	CK2_PHOSPHO_SITE	PDOC00006
PS00006	580->584	CK2_PHOSPHO_SITE	PDOC00006
PS00006	668->672	CK2_PHOSPHO_SITE	PDOC00006
PS00007	537->546	TYR_PHOSPHO_SITE	PDOC00007
PS00008	25->31	MYRISTYL	PDOC00008
PS00008	43->49	MYRISTYL	PDOC00008
PS00008	114->120	MYRISTYL	PDOC00008

PS00008	326->332	MYRISTYL	PDOC00008
PS00008	385->391	MYRISTYL	PDOC00008
PS00008	514->520	MYRISTYL	PDOC00008
PS00008	622->628	MYRISTYL	PDOC00008
PS00009	287->291	AMIDATION	PDOC00009
PS00009	436->440	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3_15g14.2)

DKFZphtes3_15h1

group: testes derived

DKFZphtes3_15h1 encodes a novel 672 amino acid protein with very weak similarity to several proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to Hsp70/Hsp90 organizing protein

complete cDNA, complete cds, no EST hits

Sequenced by GBF

Locus: unknown

Insert length: 2277 bp

Poly A stretch at pos. 2252, polyadenylation signal at pos. 2226

```
1 AAACCAGATA GAGGTTCTCC AGCTTTTCTT TGATTGTCTC TGCTTTAGCG
51 TCTCTAAATC CGGTCACCAT GTCGGACCCC GAAGGCGAGA CCTTGCGAAG
101 CACCTTTCCC TCTTATATGG CCGAAGGCGA GCGGCTCTAC CTGTGCGGGG
151 AATTTTCTAA AGCCGCGCAG AGCTTCAGCA ACGCTCTTTA CCTTCAGGAT
201 GGAGACAAGA ACTGCCTGGT TGCTCGCTCA AAGTGCTTCC TGAAGATGGG
251 AGACTTGGAG AGATCCCTGA AGGATGCTGA GGCTTCGCTC CAGAGTGACC
301 CAGCTTCTCT TAAGGGGATT TTGCAAAAGG CTGAGACACT GTACACCATG
351 GGAGACTTTG AGTTTGCTTT GGTATTCTAT CATCGAGGCT ACAAGCTGAG
401 GCCTGATCGG GAATTCAGAG TTGGCATTC AAGAGCCAG GAAGCCATCA
451 ACAACTCAGT GGAAGTCCT TCTTCCATTA AGCTGGAGAA CAAAGGGGAC
501 CTCTCCTTCT TAAGCAAGCA GGCTGAGAAT ATAAAAGCCC AGCAGAAGCC
551 TCAGCCCATC AAACACCTCT TACACCCAC CAAGGGAGAG CCCAAGTGGA
601 AGGCCTCGCT CAAGAGTGAG AAGACTGTCC GCCAGCTTCT GGGGGAGCTC
651 TACGTGGACA AAGAGTATTT GGAGAAGCTC CTATTGGATG AAGACCTGAT
701 CAAAGGCACC ATGAAGGGCG GCCTGACTGT GGAGGACCTC ATCATGACGG
751 GCATCAACTA CCTGGATACT CACAGCAACT TCTGGAGGCA GCAGAAGCCG
801 ATCTACGCCA GGGAGCGGGA CCGGAAGCTG ATGCAAGAGA AATGGCTGCG
851 GGACCAACAA CGCCGTCCCT CACAGACAGC CCATTACATC CTCAAGAGCC
901 TGGAGGACAT TGATATGTTG CTCACAAGTG GCAGTGCTGA AGGGAGTCTT
951 CAGAAAGCTG AGAAAGTGCT GAAGAAGGTA CTGGAATGGA ACAAGGAAGA
1001 GGTACCCAAC AAGGATGAAC TGGTTGGAAA CTTGTATAGC TGCATAGGGA
1051 ATGCCAGATG TGAGCTGGGG CAGATGGAGG CAGCCCTGCA GAGCCACAGA
1101 AAGGACCTGG AGATCGCCAA GGAATATGAC CTTCTGATG CAAAATCGAG
1151 AGCCCTTGAC AACATTGGCA GAGTTTTTGC CAGAGTTGGG AAATTCAGC
1201 AAGCCATTGA CACGTGGGAA GAAAAGATCC CTCTGGCAAA AACCACCTTG
1251 GAGAAGACCT GGCTGTTCCA CGAGATCGGC CGTGCTACT TGGAGCTGGA
1301 CCAGGCCTGG CAGGCCCAGA ATTATGGCGA GAAGTCCAG CAGTGTGCCG
1351 AGGAGGAAGG GGACATTGAG TGGCAACTGA ATGCCAGTGT TCTGGTGGCC
1401 CAGGCACAAG TGAAGCTGAG AGACTTCGAG TCAGCCGTGA ACAATTTTGA
1451 GAAGGCCCTG GAGAGAGCAA AGCTTGTGCA TAACAACGAG CGCGAGCAGG
1501 CCATCATCAG TGCCTTGGAC GATGCCAACA AGGGTATCAT CAGAGAAGCTG
1551 AGGAAAACCA ACTACGTGGA GAATCTCAA GAAAAAGCG AGGGAGAAGC
1601 TCACTGTAT GAAGATAGAA TAATAACAAG AGAGAAGGAC ATGAGGAGAG
1651 TGAGAGATGA GCCCGAGAAG GTGGTGAAGC AGTGGGACCA TAGTGAGGAT
1701 GAGAAAGAGA CAGATGAGGA CGATGAGGCT TTTGGGGAAG CTCTGCAGAG
1751 CCCAGCAAGC GGAAGCAGA GTGTGGAAGC AGGAAAAGCC AGAAGCGATT
1801 TGGGAGCAGT TGCCAAGGGC CTGTCAGGAG AATTAGGCAC AAGATCAGGA
1851 GAAACAGGCA GGAAGCTACT AGAAGCTGGC AGAAGAGAGT CAAGAGAAAT
1901 TTATAGGAGG CCTTCGGGAG AATTAGAGCA AAGACTCTCA GGAGAATTCA
1951 GCAGACAGGA ACCAGAAGAA CTAAGAAGAA TTTCAGAAGT GGGCAGAAGA
2001 GAGCCAGAAG AACTGGGAAA AACACAATTG GGAGAAATAG GAGAAACGAA
2051 AAAAAACAGGA AATGAGATGG AAAAGGAATA TGAATGAAGC CATCGGTAGA
2101 GATGAGGATC AGGAAGCTGG TGTTCAGAGG GATCATGGGA TTTTATTAAA
2151 CTGGATTTC AAGCGATTG TCTGTTATAG GAAAAATGAG GGTTTTACTT
2201 CTGCTGCTTT CCATCACTAT TTTGCCATTA AATAGGTGTC TTTCACTCTT
2251 GCAAAAAAAA AAAAAA AAAA
```

BLAST Results

No BLAST result

No Medline entry

Peptide information for frame 3

ORF from 69 bp to 2084 bp; peptide length: 672
Category: similarity to known protein

1	MSDPEGETLR	STFPSYMAEG	ERLYLCGEFS	KAAQSFSNAL	YLQDGDKNCL
51	VARSCKFLKM	GDLSRLSKDA	EASLDQSDPAF	CKGILQKAET	LYTMDGFFFA
101	LVFYHRYGKL	RPDREFRVGI	QKAQEAINNS	VGSPSSIKLE	NKGDSLFLSK
151	QAENIKAQQK	PQPMKHLHP	TKGEPKWKAS	LKSEKTVRQL	LGELYVQKEY
201	LEKLLLEDLD	IKGTMKSGST	VEDLIMTGIN	YLDTHSNFVR	EQKLPYIARV
251	DRKLMQEKWL	RDRHKRPSQT	AHYILSKLED	IDMLTLTSGSA	EGGSLQAEKV
301	LKKVLEWNKE	EVPNKDELVG	NLYSCIGNAQ	IELGQMEAL	QSHRKDLEIA
351	KEYDLPDASK	RALDNIQGRV	ARVGKQQAI	DTWEEKIPLA	KLTEKTLWLF
401	HEITGRXYLE	DQAWQQAQNG	EKSQQAEE	GDI EQWLNAS	VLVAQAQKVL
451	RFDESAVNNF	EKALAREAKL	HNNEAQAQII	SALDDANQGI	IRELKRNTYV
501	ENLKEKSEGE	ASLYEDRIIT	RKEDMRRVDR	EPEKVVQKWD	HSEDEKTDNE
551	DDEAFGEALQ	SPASGKQSV	AGKARSDLGA	VAKGLSGELG	TRSGETGRKL
601	LEAGRRESRE	IYRRPSEGE	QRLSGEFSRQ	EPEELKKLSE	VGRREPEELG
651	KTOFGETGET	KKTNGMEKE	YE		

BLASTP hits

Entry AF039202.1 from database TREMBL:
product: "Hsp70/Hsp90 organizing protein"; *Cricetulus griseus*
Hsp70/Hsp90 organizing protein mRNA, complete cds.
Score = 149, P = 5.3e-07, identities = 42/160, positives = 74/160

Entry AI09782_1 from database TREMBL:
product: "myosin heavy chain"; *Argopecten irradians* myosin heavy chain
mRNA, complete cds.
Score = 155, P = 6.1e-07, identities = 140/623, positives = 256/623

Entry S56658 from database PIR:
stress-induced protein stil - soybean
Score = 156, P = 9.7e-08, identities = 41/153, positives = 72/153

Alert BLASTP hits for DKFZphtes3 15h1, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3 15h1, frame 3

Report for DKFZphtes3_15h1.3

```
[LENGTH]          672
[MW]               76655.61
[pI]              5.49
[HOMOL]           PIR:S56658 stress-induced protein stil - soybean 6e-10
[SUPFAM]          tetratricopeptide repeat homology 1e-07
[PROSITE]         MYRISTYL          7
[PROSITE]         AMIDATION         3
[PROSITE]         CAMP_PHOSPHO_SITE 4
[PROSITE]         CK2_PHOSPHO_SITE  15
[PROSITE]         TYR_PHOSPHO_SITE   1
[PROSITE]         PKC_PHOSPHO_SITE   11
[PROSITE]         ASN_GLYCOSYLATION  2
[KW]              All_Alpha
[KW]              LOW COMPLEXITY      4.76 %
```

```

SEQ      MSDPEGETLKSTFPSYMAEGERLYLCGEFSKAAQFSFNALYQDGDKNCLVARSKCFLKM
SEG
PRD      cccccccceeeccccccccccccccccccccchhhhhhhhhhhhhhhccccceehhhhhhhhhhh

SEQ      GDLERSLKDAEASLQSDPAFCCKILQKAETLYTMGDFEFALVFYHRGYKLRDPREFRVGI
SEG
PRD      hcchhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccchhhhhh

```


Prosites for DKFZphtes3_15h1.3

593

PS00008	288->294	MYRISTYL	PDOC00008
PS00008	320->326	MYRISTYL	PDOC00008
PS00008	334->340	MYRISTYL	PDOC00008
PS00008	590->596	MYRISTYL	PDOC00008
PS00009	596->600	AMIDATION	PDOC00009
PS00009	603->607	AMIDATION	PDOC00009
PS00009	641->645	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3_15h1.3)

DKFZphtes3_15i5

group: cell structure and motility

DKFZphtes3_15i5 encodes a novel 717 amino acid protein with similarity to radial spokehead proteins.

The novel protein is similar to the *Chlamydomonas reinhardtii* radial spokehead protein of flagella or axoneme and to the *Strongylocentrotus purpuratus* sea urchin spermatozoa protein p63. This protein is important for the maintenance of a planar form of sperm flagellar beating. In addition, the novel protein contains a transferrin signature 1 for iron-binding. The new protein seems to be a part of the human radial spoke heads in spermatozoa.

BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in modulating the structure of the human spermatozoa radial spoke head and modulation of sperm motility in men.

strong similarity to "radial spokehead" proteins

complete cDNA, complete cds, 1 EST hit (from a testis library)
"radial spokehead" part of flagella in *Chlamydomonas*, this protein seems to be part of the sperm motor or tail

Sequenced by GBF

Locus: unknown

Insert length: 2478 bp

Poly A stretch at pos. 2452, polyadenylation signal at pos. 2433

```
1  CACCCCTGGCC CGCTCCCCGC GCCCTCCACG GGTAAACGGCC CCCTCTCTCG
51 GTGCTCAGAA ACCGGCGGTG TCGACAGGTG GCTCTCGCTT GGCCTCCTTG
101 TCTGCAAGCC TTTCTCCTAG AGATCTGTGC CTCCTGGCGA ACCATGGGAG
151 ACCTGCCGCC CTACCCTGAG CGCCCTGCCC AGCAGCCTCC GGGCCGGAGG
201 ACTTCTCAGG CCTCCCAGAG GCGGCACAGT CGGGACCAAG CTCAGGCCCT
251 GGCAGCGGAC CCCGAGGAGA GGCAGCAGAT ACCTCCAGAC GCCCAGCGAA
301 ACGCCCTTGG TTGGTCACAG AGGGGCAGCC TGTCCCAACA GGAGAACTTG
351 CTGATGCCCC AGGTCTTCCA GGCTGAGGAA GCCCGGCTGG GTGGCATGGA
401 GTACCCATCT GTGAACACGG GCTTTCCTTC AGAGTTCCAG CCTCAGCCTT
451 ACTCTGATGA AAGCAGGATG CAGGTCGCGC AGCTCACCAC CAGCCTAATG
501 CTGCAGCGGC TCCAGCAGGG CCAAAGCAGC CTGTTCCAGC AACTGGACCC
551 CACCTTCCAG GAGCCCCCAG TCAACCCCTT GGGCCAGTTC AACCTCTACC
601 AGACCAACCA GTTCTCTGAA GGTGCCCAGC ACGGGCCTTA CATAAGGGAT
651 AGACCTGCGG TTCAGTTCTT GCCCTCTGAG CTGGGCTTCC CACACTACAG
701 TGCCCCAGGTG CCTGAGCCCG AGCCTCTGGA GCTGGCCGTG CAGAACGCCA
751 AGGCCTACCT GCTGCAGACC AGCATCAATT GCGACCTCAG CCTGTACGAG
801 CACCTGGTAA ATCTGCTGAC CAAGATCCTG AACCAGCGGC CTGAGGACCC
851 CTTGTCTGTC CTGGAGTCTC TGAACCGCAC CACGCAGTGG GAGTGGTTCC
901 ACCCCAAGCT GGACACGCTG CGGGACGACC CCGAGATGCA GCCCACTAC
951 AAGATGGCGG AGAAACAGAA GCGGCTGTTC ACCCGGAGTG GAGGCGGCAC
1001 TGAAGGCGAA CAGGAGATGG AGGAGGAGGT GGGGGAGACA CCAGTGCCCA
1051 ACATCATGGA GACTGCCTTC TACTTCGAGC AGGCCGGCGT CGGCCTGAGC
1101 TCGGACGAGA GCTTCCGCAT TTTCTTGGCC ATGAAACAGC TGGTGGAGCA
1151 GCAGCCCATC CACACCTGTC GCTTCTGGGG CAAGATCCTG GGAATCAAAC
1201 GCAGCTACCT GGTGGCCGAG GTGGAATTCC GGGAGGGCGA GGAGGAGGCA
1251 GAGGAGGAGG AGGTGGAGGA GATGACGGAA GGTGGCGAGG TCATGGAGGC
1301 GCACGGCGAG GAGGAGGGCG AGGAGGACGA GGAGAAGGCC GTGGACATCG
1351 TCCCTAAGTC CGTATGGAAG CCGCCGCCCG TGATCCCAAA GGAGGAGAGC
1401 CGCTCAGGCG CCAACAAGTA CCTGTACTTT GTGTGCAACG AGCCGGGCCT
1451 GCCATGGACG CGGCTGCCCC ACGTCACTCC AGCCAGATC GTGAACGCCC
1501 GAAAGATCAA GAAGTTCTTC ACAGGCTACC TGGACACGCC AGTCGTGAGC
1551 TACCCACCCT TCCCGGGCAA CGAGGCCAAC TACCTGCGGG CCCAGATAGC
1601 CCGCATCTCG GCCGCCACGC AGGTGAGCCC GCTGGGCTTC TACCAGTTTA
1651 GTGAGGAGGA GGGCGACGAG GAGGAGGAAG GTGGTGCTGG GCGCGACTCC
1701 TACGAGGAGA ACCCGGACTT CGAGGGCATC CCCGTGCTGG AGCTGGTCTG
1751 CTCATGGGCC AACTGGGTGC ATCACACACA GCACATCCTG CCGCAGGGCC
1801 GCTGCACCTG GGTGAACCTT TTGCAGAAGA CAGAGGAGGA GGAGGACCTG
1851 GGGGAGGAGG AAGAGAAGGC AGATGAGGGG CCAGAGGAGG TGGAGCAGGA
1901 GGTGGGCCCC CCACTGTCTA CGCCACTTTC AGAAGATGCA GAAATCATGC
1951 ACCTGGCACC CTGGACCACC CGCCTGTCTT GCAGCCTCTG CCCGAGTAC
2001 TCAGTGGCCG TTGTGCGCTC CAACCTCTGG CCCGGGCCT ATGCCTATGC
2051 CAGTGGCAAA AAGTTTGAGA ACATCTACAT CGGCTGGGGT CACAAGTACA
2101 GCCCCGAGAG CTTCAACCCG GCCCTGCCAG CCCCCATCA ACAAGAGTAC
2151 CCCAGTGGCC CAGAGATCAT GGAGATGAGT GACCCACAG TGGAGAGGGA
2201 GCAGGCTCTG AAAGCAGCCC AGGAACAAGC CCTGGGAGCC ACAGAGGAGG
2251 AGGAGGAGG CGAGGAGGAG GAGGAGGGCG AGGAGACAGA TGAAGGAGG
```

2301 CCACCCTCTA GGCACCTTCC CCAAGCAGGT AGATAGCAAA TTTCCCCTTA
 2351 GAGGTAGTTA GCATGGATTA TATTTTCACT ATGTGCTTCC TGTCCTCAGA
 2401 GGGCAGGGAT AGAAAAGGAA GGCAACTGCT TCAAATAAAA TTCCTCCACG
 2451 GCATTAATAA AAAAAAAAAA AAAAAAAG

BLAST Results

No BLAST result

Medline entries

86251010:

Molecular cloning and expression of flagellar radial spoke and dynein genes of *Chlamydomona*

81142496:

Radial spokes of *Chlamydomonas* flagella: polypeptide composition and phosphorylation of stalk components.

9450971:

Molecular cloning and characterization of a radial spoke head protein of sea urchin sperm axonemes: involvement of the protein in the regulation of sperm motility.

Peptide information for frame 3

ORF from 144 bp to 2294 bp; peptide length: 717
 Category: strong similarity to known protein

1 MGDLPYPYPER PAQPPGRRT SQASQRRHSR DQAQALAADP EERQQIPPDA
 51 QRNAPGWSQR GSLSQQENLL MPQVFQAEAA RLGGMEYPSV NTGFPSEFQP
 101 QPYSDESRRM VAEITTSML QRLQQGQSSL FQQLDPTFQE PPVNPLGQFN
 151 LYQTDQFSEG AQHGPIYRDD PALQFLPSEL GFPHYSAQVP EPEPLELAVQ
 201 NAKAYLLQTS INCDSLSEYH LVNLLTKILN QRPEDPLSVL ESLNRTTQWE
 251 WFHPKLDTLR DDPQMPTYK MAEKQKALFT RSGGGTEGEQ EMEEEVGETP
 301 VPNIMETAFY FEQAGVGLSS DESFRIFLAM KQLVEQQPIH TCRFWGKILG
 351 IKRSYLVAEV EFREGEAEAE EEEVEEMTEG GEVMEAHGEE EGEEDEEKAV
 401 DIVPKSVWKP PPVIPKEESR SGANKLYFV CNEPGLPWTR LPHVTPAQIV
 451 NARKIKKFFT GYLDTPVVS Y PPFPGNEANY LRAQIARISA ATQVSPGLGFY
 501 QFSEEEGDEE EEGGAGRDSY EENPDFEGIP VLELVDSMAN WVHHTQHILP
 551 QGRCTWVNPL QKTEEEEDLG EEEKKADEGP EEVEQEVGPP LLTPLSEDAE
 601 IMHLPWTTR LSCSLCPQYS VAVVRSNLWP GAYAYASGKK FENIYIGWGH
 651 KYSPESEFNP LPAPIQYEY SGPEIMEMSD PTVEEQALK AAQEALGAT
 701 EEEEGEETDD EGEETDD

BLASTP hits

Entry U73123_1 from database TREMBL:

product: "radial spokehead"; *Strongylocentrotus purpuratus* radial spokehead mRNA, complete cds.

Score = 1604, P = 7.4e-165, identities = 303/523, positives = 395/523

Entry B44498 from database PIR:

radial spoke protein 6 - *Chlamydomonas reinhardtii*

Score = 386, P = 3.4e-45, identities = 105/264, positives = 138/264

Alert BLASTP hits for DKFZphtes3_15i5, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_15i5, frame 3

Report for DKFZphtes3_15i5.3

[LENGTH] 717
 [MW] 80913.61
 [pI] 4.36

[HOMOL] TREMBL:U73123_1 product: "radial spokehead"; Strongylocentrotus purpuratus
radial spokehead mRNA, complete cds. 1e-130
[PROSITE] TRANSFERRIN_1 1
[PROSITE] MYRISTYL 5
[PROSITE] AMIDATION 2
[PROSITE] CAMP_PHOSPHO_SITE 2
[PROSITE] CK2_PHOSPHO_SITE 14
[PROSITE] TYR_PHOSPHO_SITE 1
[PROSITE] GLYCOSAMINOGLYCAN 1
[PROSITE] PKC_PHOSPHO_SITE 8
[PROSITE] ASN_GLYCOSYLATION 1
[KW] All Alpha
[KW] LOW_COMPLEXITY 21.48 %

SEQ MGDLPYPYPERPAQPPGRRTSQASQRRHSRDQAQALAADPEERQIIPDAQRNAPGWSQR
SEGXXXXXXXXXX
PRD ccc

SEQ GSLSQQENLLMPQVFAEEEARLGGMEYPSVNTGFPSEFQPPYSDSRMQVAELTSLML
SEGXXXXXXXXXX
PRD ccc

SEQ QRLQQGSSSLFQQLDPTFQEPVPLGQFNLYQTDQFSEGAQHGPYIRDDPALQFLPSEL
SEGXXXXXXXXXX
PRD hhhhhcc

SEQ GFPHYSAQVPEPELELAVQNAKAYLLQTSINCDLSLYEHLVNLTKILNQRPEPLSVL
SEGXXXXXXXXXX
PRD ccc

SEQ ESLNRTTQWEWFHPKLDLRLDDPEMQPTYKMAEKQKALFTRSGGGTEGEQEMEEVGETP
SEGXXXXXXXXXX
PRD hhhchhhhhcc

SEQ VPNIMETAFYFEQAGVGLSSDESFRIFLANKQLVEQQPIHTRFWGKILGKRSYLVAEV
SEGXXXXXXXXXX
PRD ccc

SEQ EFREGEAEAEVEEMTEGGEVMEAHGEEGEEDEEKAVDIVPKSVWKPPVPIKEESR
SEGXXXXXXXXXX
PRD hhhhhhhhhhhhhhhhhhhhhcc

SEQ SGANKYLYFVCNEPLPWTRLPHVTPAQIVNARKIKKFFTGYLDPVVSYPFFPGNEANY
SEGXXXXXXXXXX
PRD ccc

SEQ LRAQIARISAATQVSPLGFYQFSEEGDEEEEGAGRDSYEENPDFEGIPVLELVDSMAN
SEGXXXXXXXXXX
PRD hhhhhhhhhhhhhhhhhhhhhcc

SEQ WVHHTQHILPQGRCTWVNPLQKTEEEEDLGEDEEKADEGPVEVEQVGPPLLTPLSEDAE
SEGXXXXXXXXXX
PRD hhhcc

SEQ IMHLAPWTRLSCSLCPQYSVAVVRNLWPGAYAYASGKKFENIYIGWGHKYSPEFNPFA
SEGXXXXXXXXXX
PRD ccc

SEQ LPAPIQEQYPSGPEIMEMSDPTVEEQALKAAQEALGATEEEEEEGEEEEGEETDD
SEGXXXXXXXXXX
PRD ccc

Prosite for DKFZphtes3_15i5.3

PS00001	244->248	ASN_GLYCOSYLATION	PDOC00001
PS00002	282->286	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	18->22	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	26->30	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	24->27	PKC_PHOSPHO_SITE	PDOC00005
PS00005	58->61	PKC_PHOSPHO_SITE	PDOC00005
PS00005	258->261	PKC_PHOSPHO_SITE	PDOC00005
PS00005	268->271	PKC_PHOSPHO_SITE	PDOC00005
PS00005	323->326	PKC_PHOSPHO_SITE	PDOC00005
PS00005	341->344	PKC_PHOSPHO_SITE	PDOC00005
PS00005	608->611	PKC_PHOSPHO_SITE	PDOC00005
PS00005	637->640	PKC_PHOSPHO_SITE	PDOC00005
PS00006	64->68	CK2_PHOSPHO_SITE	PDOC00006
PS00006	137->141	CK2_PHOSPHO_SITE	PDOC00006

PS00006	216->220	CK2_PHOSPHO_SITE	PDOC00006
PS00006	238->242	CK2_PHOSPHO_SITE	PDOC00006
PS00006	247->251	CK2_PHOSPHO_SITE	PDOC00006
PS00006	258->262	CK2_PHOSPHO_SITE	PDOC00006
PS00006	286->290	CK2_PHOSPHO_SITE	PDOC00006
PS00006	319->323	CK2_PHOSPHO_SITE	PDOC00006
PS00006	503->507	CK2_PHOSPHO_SITE	PDOC00006
PS00006	519->523	CK2_PHOSPHO_SITE	PDOC00006
PS00006	563->567	CK2_PHOSPHO_SITE	PDOC00006
PS00006	671->675	CK2_PHOSPHO_SITE	PDOC00006
PS00006	682->686	CK2_PHOSPHO_SITE	PDOC00006
PS00006	700->704	CK2_PHOSPHO_SITE	PDOC00006
PS00007	639->646	TYR_PHOSPHO_SITE	PDOC00007
PS00008	284->290	MYRISTYL	PDOC00008
PS00008	315->321	MYRISTYL	PDOC00008
PS00008	350->356	MYRISTYL	PDOC00008
PS00008	435->441	MYRISTYL	PDOC00008
PS00008	475->481	MYRISTYL	PDOC00008
PS00009	16->20	AMIDATION	PDOC00009
PS00009	637->641	AMIDATION	PDOC00009
PS00205	619->628	TRANSFERRIN_1	PDOC00182

(No Pfam data available for DKF2phtes3_15i5.3)

DKFZphtes3_15j18

group: testes derived

DKFZphtes3_15j18 encodes a novel 148 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, few EST hits

Sequenced by GBF

Locus: unknown

Insert length: 905 bp

Poly A stretch at pos. 839, polyadenylation signal at pos. 815

```
1 GTGATTCATA TGCTTCCATA GCAGGTGTCT GCTTCTGAGC CAAGCTCCCA
51 GGGCAGCGGA GCAGGCACCA ACCAGCATCC CAGGGGAGGG CACAGCTTGT
101 CCAGCTGGGA TGTGTGGGTG CCCTGTGAGA TGCCCCAAGC CACCAACCCA
151 GCTTATCTCA GGAGAAGCCT CGGCGGCCCG TCTGCCGGCC TGGAGAGATG
201 TGCTACAGCA GCCGGGGGTG GGGGGAGAGG GTGGGCTTAG AATCTCTTGG
251 CAGGGAGCCC CCAAGAGCAG GGTGAGACCT GCCTTCATTT CACCTGTCCC
301 CTTACAGTT CTGCAAAGCC AGCATTATCA TCCCTTTTCA GAAGGAGTGG
351 GCACTCAGGT GGAATGCCTC ACCCCAGTCC TCGGGCTGGA AAGCGATATG
401 GCCAGGACTG CACCCACCCC CTCATCCCTG CACCCCTTCC CTGCCTGGGA
451 TTCCTCCAGC CCTGTGCACT GTGGAGCGCC TCTGCCTTCC GCTCATGGAG
501 GTTCCCAAG GGCACGCGCT GAGGGCAGCT GGTCTCAGCC TGGGGCCGGG
551 TCCTAGTAAC TGTCTCTCTT TGCTTTCCAG CCAGTGTTTT GGGGTTTGAA
601 GTTGAATCT TCAGTACTG TCAAGAACAG CCACAAAAAT GTGTCACGAT
651 CAAGATCTTT GAGAGTCCAC CAATCAGGAG GCGTCTGTGA CAGTCGCTGT
701 CTTCTCAGAA CAGAATCCAC ACCCAGGATT CAACCCAAAT GATTTCAT
751 CAGGTGATTC TTGGTTGTAG CAAAGTTCAT GTGAATGTGG GTGAGTTTCT
801 GTTATGAATG TGGTCAATAA ATGTTATTG TGAACCTCTA AAAAAAAAAA
851 AAAAAAAAAA GCGGCCGCT CTAGAGGATC CAAGCTTACG TACGCGAAAA
901 AAAAG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 110 bp to 553 bp; peptide length: 148
Category: putative protein

```
1 MFGCPVRCPK PPTQLISGEA SAARLPWARD VLQPGVGGG GGLRISWQGA
51 PKSRVRPAFI SPVPFTVLQS QHYHPFSEGV GTQVECLTPV LRLESDMART
101 APHPSSLHPF PAWDSSSPVH CGAPLPSAHG GFPRARAEGS WSQPGAGS
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_15j18, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphtes3_15j18, frame 2

Report for DKFZphtes3_15j18.2

[LENGTH] 148
[MW] 15665.78
[pI] 8.91
[PROSITE] MYRISTYL 3
[PROSITE] CK2_PHOSPHO_SITE 1
[KW] Irregular

SEQ MFGCPVRCPKPPTQLISGEASAARLPAWRDVLQQPGVGEGGLRISWQGAPKSRVRPAFI
PRD cccccccccccccccccccccchhhhhhccccccccceeeeecccccccccccccc

SEQ SPVPFTVLQSQHYHPFSEGVGTQVECLTPVLRLESDMARTAPHPSSLHPFPWDSSSPVH
PRD cccccceccccccccccccccccchhhhhhcccccccccccccccccccccccc

SEQ CGAPLPSAHGGFPRARAEGSWSQPGAGS
PRD ccccccccccccccccccccccccc

Prosites for DKFZphtes3_15j18.2

PS00006	82->86	CK2_PHOSPHO_SITE	PDOC00006
PS00008	38->44	MYRISTYL	PDOC00008
PS00008	42->48	MYRISTYL	PDOC00008
PS00008	49->55	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_15j18.2)

DKFZphtes3_15j3

group: nucleic acid management

DKFZphtes3_15j3 encodes a novel 743 amino acid protein with similarity to proteins with unknown function.

The novel protein contains a RNA recognition motif, predicted by Pfam and therefore binds to RNA. The protein is similar to YGR276c, a ribonuclease H of *S. cerevisiae*. Thus, the protein seems to a new RNA-modifying protein.

The new protein can find application in modulating the RNA metabolism in human cells and as a tool for biotechnologic manipulations.

"44M2.3"; product, differences to genmodel, similarity to ribonuclease H

complete cDNA, complete cds, EST hits
YGR276c = ribonuclease H
differences to genmodel of 44M2.3

Sequenced by GBF

Locus: /map="16p11.2"

Insert length: 2695 bp

Poly A stretch at pos. 2601, polyadenylation signal at pos. 2579

```

1 GCGGTTGTTG TTGGCAGCTG TGGCTAAGGA GGGGAGAACC TCTGCTCCCC
51 GCCCGTCTTC TCTTCTGCGT TTCCCGGGCT AGGGGGCGTG GGGAGTGGTT
101 TTAGGCGGCG AAGCCGCTCG GCAGCACCTT CTTCTTTGC CAGGCAGACG
151 CCCGTTGTAG CCGTTGGGGA ACCGTTGAGA ATCCGCCATG GAGCCAGAGA
201 GGGAAAGGGAC CGAGAGACAC CCCAGGAAGG TCAGGGAAAG CAGGCAGGCC
251 CCAATAAAGC TGGTCGGGGC AGCTGAGGCG ATGAAAGCCG GTTGGGATCT
301 CGAGGAGAGT CAGCCCGAGG CCAAGAAAGC CCGCTTATCT ACCATTTTAT
351 TTAAGCATTC GACATAAAT CCGCTTGCCG AAACCCAGCT GGTGCCAGCT
401 TATGCAATTC TGGGCAAATC CAATGTTCCA AAACCCAGCT GGTGCCAGCT
451 TTTTCATCAA AACACCTAA ACAACGTAGT GGTTTTGTCT CTGCAGGGAA
501 TGAGTCAGCT ACACCTTTAC AGGTTCTATT TGGAGTTTGG ATGCTTTCGA
551 AAAGCATTCG GACATAAAT CCGCTTGCCG CCACCATCAT CTGATTTTCT
601 AGCTGATGTT GTTGGGCTAC AAAGCTGAAC AAGAGCTGGA GATCTGCCCA
651 AGACAAATGA AGGGCCCTTA CTTCTTAATG CAAAAGCCGC CATCAACCTT
701 CAGGAATGAT CCATCATTCA AAAGTATGGC TCTAAGAAAG TGGGCTTGAC
751 CAGATGCCCT CTGACAAAGG AGGAAATGAG AACGTTTCAC TTTCCATTAC
801 AAGGTTTTCC TGATTGTGAA AACTTTTTAC TTACCAATAT TAATGGTTCT
851 ATAGCAGACA ATAGTCTCTT CTTGGACTT GACTGTGAAA TGTGCTTCAC
901 ATCCAAGGGG AGAGAGCTAA CACGCATCTC ACTGGTTGCT GAAGGAGGCT
951 GCTGTGTTAT GGATGAACCT GTCAAACCTG AAAACAAGAT TCTGGACTAC
1001 CTCACCAAGT TTTGGGGAAT CACGAAGAAG ATTCTTAACC CAGTGACGAC
1051 CAAACTCAAA GATGTACAGA GGCAGTAAA AGCACTGCTT CCTCCTGATG
1101 CTGTGTGTTG GGGCCACTCC TTAGATTGGG ATCTCAGAGC ACTGAAATG
1151 ATACATCCAT ATGTTATTGA TACATCGTTG CTTTATGTCA GAGAGCAGGG
1201 CAGAAATGAT AAGCTCAAGT TCTTAGCCAA AGTTATTTTG GGGAAAGGATA
1251 TACAGTGTCC AGACAGACTT GGTCAATGAT CCACAGAAGA TGCTAGAACA
1301 ATCCTTGAAT TGGCTCGGTA TTTCTTAAG CATGGCCCAA AAAAGATTGC
1351 AGAAGTAAAT CTAGAAGCAC TAGCTAATCA CCAAGAAATA CAAGCAGCAG
1401 GCCAAGAGCC TAAAAACACA GCAGAAGTAC TTCAGCACCC AAACACAAGT
1451 GTTTTAGAAT GCTTGGATTC AGTGGGTGAG AAGCTTCTTT TTTTGACCCG
1501 GGAGACAGAT GCTGGTGAAC TTCCATCTTC CAGAAATGTT CAAACTATTA
1551 AGTGTCTTTC AAATAAAGAG GTTCTTGAGC AGGCCAGAGT GGAATCCCC
1601 CTGTTTCCCT TCAGCATTGT TCAGTTCTCT TTTAAGGCCT TTTACCTGT
1651 CCTCACTGAG GAGATGAACA AAAGGATGAG GATCAAGTGG ACAGAGATAT
1701 CAACTGTCTA TGCTGGGCCA TTAGCAAAA ATTGCAATCT CAGGGCTCTG
1751 AAGAGGCTGT TTTAAAGCTT TGGCCAGTCC CAGTCAATGA CTTTGTGTCT
1801 TGAAACCCGT CAGGTGCAGA GGCCTGTGAC AGAGCTCAGC CTTGATTGTG
1851 ACACCCCTCGT GAATGAGCTG GAAGGAGATT CTGAAAACCA AGGCTCTATA
1901 TATCTGTCTG GAGTGAGTGA AACCTTCAAA GAACAGCTAT TGCAGGAGCC
1951 CCGCCTCTTT CTTGGCCTGG AAGCTGTGAT CTTGCCTAAA GATCTTAAA
2001 GTGGAAAGCA GAAAAAATAC TGTTCCTGTA AATTCAAAAG TTTTGGCAGT
2051 GCCCAGCAGG CCCTCAACAT TCTCAGAGG AAGGACTGGA AGCTGAAAGG
2101 CAGGCATGCC CTAACCCCCA GGCACCTCCA TGCTGGGCTC AGAGGCTTAC
2151 CACCTGAATC AACCAAGGCTC CCAGGGCTTC GTGTTGTACC TCCCCCTTT
2201 GAACAGGAGG CTTTGACAGC TGTGAACTG GACCACCCGA AGATAGCAGC
2251 CTGGCCGCTG AGCCGGAAGA TTGGAAAGCT CTACAACAGC TTGTGCCCGG
2301 GCACTCTCTG CCTCATCCTG CTGCCAGGAA CCAAGAGCAC TCATGGTTCA
2351 CTCTCTGGTC TAGGACTGAT GGAATAAAA GAGGAAGAAG AAAGCGCTGG
2401 CCCAGGCCTG TGTTCTGTAG TCGGCCTGCC ATGTTTCCAT GTGCCATTTT

```

```

2451 TTACCCCTTG TAGGCAATGG CAAAGAATGT GGTGAGGCTG TAGCCTCCCC
2501 AACCCAGCAGA CAGTTTATG GAACTTGGT ATAGCAGCTA AAAGAGTTTA
2551 GTTTGTTTAT ATGGCATGTA TAAGTTTCA ATAAATGCCT AAAGTTCAAG
2601 CATAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2651 AGGGCGGCCG CTCTAAAGGA TCCAAGCTTA CGTACGCGAA AAAAG

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 188 bp to 2416 bp; peptide length: 743
 Category: similarity to known protein

```

1  MEPEREGTER HPRKVRESRQ APNKLVGAAE AMKAGWDL EE SQPEAKKARL
51  STILFTDNCE VTHDQLCELL KYAVLGKSNV PKPSWCQLFH QNHLNNVVVF
101 VLQGMSQLHF YRFYLEFGCL RKAFRHKFRL PPPSSDFLAD VVGLQTEQRA
151 GDLPKTMEGP LPSNAKAAN LQDDPIIQKY GSKKVGLTRC LLTKEEMRTF
201 HFPLQGFPDC ENFLLTKCNG SIADNSPLFG LDCMCLTSK GRELTRISLV
251 AEGGCCVMDE LVKPENKILD YLTSFSGITK KILNPVTTKL KDVQRQLKAL
301 LPPDAVLVGH SLDLRLALK MIHPYVIDTS LLYVREQRRR FKLKFLAKVI
351 LGKDIQCPDR LGHDATEDAR TILELARYFL KHGPKKIAEL NLEALANHQE
401 IQAAGQEPKN TAEVLQHPNT SVLECLDSVG QKLLFLTRET DAGELPSSRN
451 CQTIKCLSNK EVLEQARVEI PLFFPSIVQF SFKAFSPVLT EEMNKRMRK
501 WTEISTVYAG PFSKNCNLRA LKRLFKSFGP VQSMTFVLET RQVQRPVTEL
551 TLDCOTLVNE LEGDSENQGS IYLSGVSETF KEQLLQEPRL FLGLEAVILP
601 KDLKSGKQKK YCFLKFKSFG SAQALNILT GKDWKLGRH ALTPRHLHAW
651 LRGLPPESTR LPGLRVVPPP FEQEALQTLK LDHPKIAAWR WSRKIGKLYN
701 SLCPGTLCIL LLPGKSTHSG SLSGGLMGI KEEESAGPG LCS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_15j3, frame 2

TREMBL:AC004381_4 gene: "44M2.3"; product: "Unknown gene product";
 Homo sapiens Chromosome 16 BAC clone CIT987SK-44M2, complete sequence.,
 N = 2, Score = 1827, P = 2.1e-284

TREMBL:AF016430_4 gene: "C05C8.5"; Caenorhabditis elegans cosmid
 C05C8., N = 2, Score = 370, P = 1.7e-34

PIR:S64609 hypothetical protein YGR276c - yeast (Saccharomyces
 cerevisiae), N = 2, Score = 334, P = 1.8e-27

TREMBLNEW:SPAC637_9 gene: "SPAC637.09"; product: "putative
 exonuclease"; S.pombe chromosome I cosmid c637., N = 3, Score = 326, P
 = 2.8e-27

>TREMBL:AC004381_4 gene: "44M2.3"; product: "Unknown gene product"; Homo
 sapiens Chromosome 16 BAC clone CIT987SK-44M2, complete sequence.
 Length = 547

HSPs:

Score = 1827 (274.1 bits), Expect = 2.1e-284, Sum P(2) = 2.1e-284
 Identities = 358/373 (95%), Positives = 358/373 (95%)

```

Query: 105 MSQLHFYRFYLEFGCLRKAFRHKFRLPPPSSDFLADVGLQTEQRA GDLPKTMEGPLPSN 164
      MSQLHFYRFYLEFGCLRKAFRHKFRLPPPSSDFLADVGLQTEQRA GDLPKTMEGPLPSN
Sbjct: 1 MSQLHFYRFYLEFGCLRKAFRHKFRLPPPSSDFLADVGLQTEQRA GDLPKTMEGPLPSN 60

Query: 165 AKAAINLQDDPIIQKYGSKKVGLTRCLLTKEEMRTFHFPLQGFPCENFLLTKCNGSIAD 224
      AKAAINLQDDPIIQKYGSKKVGLTRCLLTKEEMRTFHFPLQGFPCENFLLTKCNGSIAD
Sbjct: 61 AKAAINLQDDPIIQKYGSKKVGLTRCLLTKEEMRTFHFPLQGFPCENFLLTKCNGSIAD 120

```

Query:	225	NSPLFGLDCEM-----CLTSKGRELTRISLVAEGGCCVMDELVKPENKIL	269
		NSPLFGLDCEM CLTSKGRELTRISLVAEGGCCVMDELVKPENKIL	
Sbjct:	121	NSPLFGLDCEMARTTTFNFSIGVLQAECLTSKGRELTRISLVAEGGCCVMDELVKPENKIL	180
Query:	270	DYLTSFSGITKKILNPVTTKLKDVRQLKALLPPDAVLVGHSLDLDRALKMIHPYVIDT	329
		DYLTSFSGITKKILNPVTTKLKDVRQLKALLPPDAVLVGHSLDLDRALKMIHPYVIDT	
Sbjct:	181	DYLTSFSGITKKILNPVTTKLKDVRQLKALLPPDAVLVGHSLDLDRALKMIHPYVIDT	240
Query:	330	SLLYVREQRRFKLKFLAKVILGKDIQCPRDLGHDATEDARTILELARYFLKHGPCKIAE	389
		SLLYVREQRRFKLKFLAKVILGKDIQCPRDLGHDATEDARTILELARYFLKHGPCKIAE	
Sbjct:	241	SLLYVREQRRFKLKFLAKVILGKDIQCPRDLGHDATEDARTILELARYFLKHGPCKIAE	300
Query:	390	LNLEALANHQEIQAAQGEPKNTAEVLQHPNTSVLECLDSVGQKLLFLTRETDA GELPSSR	449
		LNLEALANHQEIQAAQGEPKNTAEVLQHPNTSVLECLDSVGQKLLFLTRETDA GELPSSR	
Sbjct:	301	LNLEALANHQEIQAAQGEPKNTAEVLQHPNTSVLECLDSVGQKLLFLTRETDA GELPSSR	360
Query:	450	NCQTIKCLSNKEV	462
		NCQTIKCLSNKEV	
Sbjct:	361	NCQTIKCLSNKEV	373

Score = 929 (139.4 bits), Expect = 2.1e-284, Sum P(2) = 2.1e-284
Identities = 175/179 (97%), Positives = 177/179 (98%)

```

Query:      538 LETRQVQRPVTELTLDGCDTLVNELEGGDSNQGSIIYLSGVSETFFKEQLLQEPRLFGLGLEAV 597
              L  ++VQRPVTELTLDGCDTLVNELEGGDSNQGSIIYLSGVSETFFKEQLLQEPRLFGLGLEAV
Sbjct:     368 LSNKEGRQPVTELTLDGCDTLVNELEGGDSNQGSIIYLSGVSETFFKEQLLQEPRLFGLGLEAV 427

Query:      598 ILPKDLKSGKQKKYCFLKFKSFGSAQQALNILTGDWKLKGRHALTPRHLHAWLRGLPPE 657
              ILPKDLKSGKQKKYCFLKFKSFGSAQQALNILTGDWKLKGRHALTPRHLHAWLRGLPPE
Sbjct:     428 ILPKDLKSGKQKKYCFLKFKSFGSAQQALNILTGDWKLKGRHALTPRHLHAWLRGLPPE 487

Query:      658 STRLPGLRVVPPPFQEALQTLKLDHPKIAAWRWSRKIGKLYNSLCPGTLCLILLPGTK 716
              STRLPGLRVVPPPFQEALQTLKLDHPKIAAWRWSRKIGKLYNSLCPGTLCLILLPGTK
Sbjct:     488 STRLPGLRVVPPPFQEALQTLKLDHPKIAAWRWSRKIGKLYNSLCPGTLCLILLPGTK 546

```

Pedant information for DKFZphtes3_15j3, frame 2

Report for DKFZphtes3_15j3.2

```
[LENGTH]          743
[MW]               83536.58
[pI]               8.87
[HOMOL]           TREMBL:AC004381_4 gene: "44M2.3"; product: "Unknown gene product"; Homo sapiens
Chromosome 16 BAC clone CIT987SK-44M2, complete sequence. 0.0
[FUNCAT]          01.03.16 polynucleotide degradation [S. cerevisiae, YGR276c] 4e-30
[FUNCAT]          99 unclassified proteins [S. cerevisiae, YLR107w] 3e-13
[FUNCAT]          05.04 translation (initiation, elongation and termination) [S. cerevisiae,
YGL094c] 1e-10
[FUNCAT]          04.05.05 mrna processing (5'-end, 3'-end processing and mrna degradation) [S.
cerevisiae, YGL094c] 1e-10
[FUNCAT]          03.22 cell cycle control and mitosis [S. cerevisiae, YOL080c] 2e-10
[PROSITE]         MYRISTYL 5
[PROSITE]         AMIDATION 1
[PROSITE]         CK2_PHOSPHO_SITE 8
[PROSITE]         TYR_PHOSPHO_SITE 1
[PROSITE]         GLYCOSAMINOGLYCAN 1
[PROSITE]         PKC_PHOSPHO_SITE 16
[PROSITE]         ASN_GLYCOSYLATION 2
[PFAM]            RNA_recognition_motif. (aka RRM, RBD, or RNP domain)
[KW]              Alpha Beta
```

SEQ PRD	MEPEREGTERHPRKRVRESQAPNKLVGAAEAMKAGWDLEESQPEAKKARLSTILFTDNCE ccchhhhhccccchhhhhhhhhcchhhhhhhhhhhccccccccccchhhhhccccccccc
SEQ PRD	VTHDQCELLKYAVLKGKSNVPKPSWCQLFHQNHLNNVVVFVLQGMSQLHFYFYLEFGGL eehhhhhhhhhhhhhhccccccccceeeccccccccceeeeeeecchhhhhhhhhhhhhhh
SEQ PRD	RKAFRHKFRLPPSSDFLADVVLGTEQRAGDLPKTMEGLPSNAKAAINLQDDPI IQKY hhhhhhhhccccccccchhhhhhhhhhhhhccccccccccccccchhhhhhhhhccccccc
SEQ PRD	GSKKVGLTRCLLTKEEMRTFHEFPLQGFPDCENFLLTCKNGSIADNSPLFGLDCEMCLTSK ccccccchhhhhhhhhhhhhhhccccccccccceeeccccccccccceeecccccccccc
SEQ PRD	GRELTRISLVAEGGCCVMDELVPKNIKLDYLTFSFGITKKILNPVTTKLKDQVQRQLKAL cchhhhhheeeccccceeeeeeccccccccccccccccccccccccchhhhhhhhhhhhh

```

SEQ      LPPDAVLVGHSLDLDRALKMHPYVIDTSLLYVREQGRFRFLKFLAKVILGKDIQCPDR
PRD      hccccccccccccchhhhhhhhhhhccccccccccccccccchhhhhhhhhhhhhhhhhcccccc

SEQ      LGHDATEDARTILELARYFLKHGPKKIAELNLEALANHQEIQAAGQEPKNTAEVLQHPNT
PRD      cccccchhhhhhhhhhhhhhhhhccccccccccccccccccccchhhhhhhhhhhhhhhhhcccccccccccccccc

SEQ      SVLECLDSVGQKLLFLTRETDALEPSSRNCQTIKCLSNKEVLEQARVEIPLFPFSIVQF
PRD      ceeeeeeccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhcccccccccccc

SEQ      SFKAFSPVLTEEMNKRMIKWEISTVYAGPFSKNCNLRALRFLKFSFGPVQSMFTVLET
PRD      eeceeeehhhhhhhhhhhhhhhheeeeeeccccccccchhhhhhhhhhhhhhhccccccccceehhhhhh

SEQ      RQVQRPVTELTLDCTLVNELEGDSENQGSIIYLSGVSETFKEQLLQEPRLFGLAEVILP
PRD      cccccccccccccchhhhhhhccccccccccccccccccccchhhhhhhhhhhhhhhcccccccccccc

SEQ      KDLKSGKQKQKCYFLFKFSFGSAQALNLTGKDWLKGRHALTPRHLHAWLRGLPPESTR
PRD      cccccccccccccccccccccccccchhhhhhhhhhhhhhhccccccccccccccccchhhhhhhcccccccccc

SEQ      LPGLRVVPPPFQEALQTLKLDHPKIAAWRWSRKIGKLYNSLCPGTLCLILPPTKSTHG
PRD      cccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhheeecccccccccccccccccccccc

SEQ      SLSGGLMGKIEEESAGPGLCS
PRD      cccccccccchhhhhhhcccccccccc

```

Prosite for DKFZphtes3_15j3.2

PS000001	219->223	ASN_GLYCOSYLATION	PDOC000001
PS000001	419->423	ASN_GLYCOSYLATION	PDOC000001
PS000002	723->727	GLYCOSAMINOGLYCAN	PDOC000002
PS000005	8->11	PKC_PHOSPHO_SITE	PDOC000005
PS000005	182->185	PKC_PHOSPHO_SITE	PDOC000005
PS000005	238->241	PKC_PHOSPHO_SITE	PDOC000005
PS000005	279->282	PKC_PHOSPHO_SITE	PDOC000005
PS000005	287->290	PKC_PHOSPHO_SITE	PDOC000005
PS000005	447->450	PKC_PHOSPHO_SITE	PDOC000005
PS000005	453->456	PKC_PHOSPHO_SITE	PDOC000005
PS000005	458->461	PKC_PHOSPHO_SITE	PDOC000005
PS000005	481->484	PKC_PHOSPHO_SITE	PDOC000005
PS000005	579->582	PKC_PHOSPHO_SITE	PDOC000005
PS000005	605->608	PKC_PHOSPHO_SITE	PDOC000005
PS000005	630->633	PKC_PHOSPHO_SITE	PDOC000005
PS000005	643->646	PKC_PHOSPHO_SITE	PDOC000005
PS000005	658->661	PKC_PHOSPHO_SITE	PDOC000005
PS000005	678->681	PKC_PHOSPHO_SITE	PDOC000005
PS000005	692->695	PKC_PHOSPHO_SITE	PDOC000005
PS000006	41->45	CK2_PHOSPHO_SITE	PDOC000006
PS000006	193->197	CK2_PHOSPHO_SITE	PDOC000006
PS000006	221->225	CK2_PHOSPHO_SITE	PDOC000006
PS000006	371->375	CK2_PHOSPHO_SITE	PDOC000006
PS000006	421->425	CK2_PHOSPHO_SITE	PDOC000006
PS000006	458->462	CK2_PHOSPHO_SITE	PDOC000006
PS000006	579->583	CK2_PHOSPHO_SITE	PDOC000006
PS000006	630->634	CK2_PHOSPHO_SITE	PDOC000006
PS000007	370->379	TYR_PHOSPHO_SITE	PDOC000007
PS000008	27->33	MYRISTYL	PDOC000008
PS000008	186->192	MYRISTYL	PDOC000008
PS000008	575->581	MYRISTYL	PDOC000008
PS000008	714->720	MYRISTYL	PDOC000008
PS000008	720->726	MYRISTYL	PDOC000008
PS000009	337->341	AMIDATION	PDOC000009

Pfam for DKFZphtes3_15j3.2

HMM_NAME	RNA recognition motif. (aka RRM, RBD, or RNP domain)		
HMM	*IYVGNLPWDtTEEDrLdLfSvQFGpIvsIrMMrDreTGRSRGFAFVEFED		
	IY+ +++ +T +E+L + + F + + + ++D G+ + ++F +F++		
Query	571	IYLSGVs-ETfKEQLLQEPRLfLGLAEAVILPKDLKSGKQKkYcFLKfKS	618
HMM	EEDaekAIdemNG..meFmGRrIRv*		
	+A+ A+ + G ++ GR +		
Query	619	FGSAQQALNIlTGKdWKLKGhALt	643

DKFZphtes3_15k11

group: signal transduction

DKFZphtes3_15k11 encodes a novel 958 amino acid protein C-terminal identical with human KIAA0781 protein and high similarity to protein kinases.

The novel protein contains a protein kinase ATP-binding region signature and a serine/threonine protein kinase active-site signature. The related murine kinase was cloned from the myocardium of the developing heart.

The new protein can find application in modulation of intracellular signal pathways dependent on this kinase.

KIAA0781, 5' extension

complete cDNA, complete cds, potential start at Bp 97, EST hits

Sequenced by GBF

Locus: /map="11"

Insert length: 4868 bp

Poly A stretch at pos. 4798, polyadenylation signal at pos. 4776

```

1 GAGCAAGCGG AGCGGCCGTC GCCCAAGCCA AGCCGCGCTG CCAACCCTCC
51 CGCCCGCCCG CGCTCCTGTC CGCCGTGTCT AGCAGCGGGG CCCAGCATGG
101 TCATGGCGGA TGGCCCGAGG CACTTGCAGC GCGGGCCGGT CCGGGTGGGG
151 TTCTACGACA TCGAGGGCAC GCTGGGCAAG GGCAACTTCG CTGTGGTGAA
201 GCTGGGGCGG CACCGGATCA CCAAGACGGA GGTGGCAATA AAAATAATCG
251 ATAAGTCTCA GCTGGATGCA GTGAACCTTG AGAAAACTTA CCGAGAAGTA
301 CAAATAATGA AAATGTTAGA CCACCCTCAC ATAATCAAAC TTTATCAGGT
351 AATGGAGACC AAAAGTATGT TGTACCTTGT GACAGAATAT GCCAAAAATG
401 GAGAAATTTT TGAATATCTT GCTAATCATG GCCGGTTAAA TGAGTCTGAA
451 GCCAGGCGAA AATTCTGGCA AATCCTGTCT GCTGTTGATT ATTGTCATGG
501 TCGGAAGATT GTGCACCGTG ACCTCAAAGC TGAATATCTC CTGCTGGATA
551 ACAACATGAA TATCAAAATA GCAGATTTCG GTTTTGGAAT TTTCTTTAAA
601 AGTGGTGAAC TGCTGGCAAC ATGGTGTGGC AGCCCCCTT ATGCAGCCCC
651 AGAAGTCTTT GAAGGGCAGC AGTATGAAGG ACCACAGCTG GACATCTGGA
701 GTATGGGAGT TGTCTTTTAT GTCTTGTCTT GTGGAGCTCT GCCCTTTGAT
751 GGACCGGACT TTCCAATTTT GAGGCAGAGG GTTCTGGAAG GAAGATTCCG
801 GATTCCGTAT TTCATGTCAG AAGATTGCGA GCACCTTATC CGAAGGATGT
851 TGGTCTTAGA CCCATCCAAA CGGCTAACCA TAGCCCAATC CAAGGAGCAT
901 AAATGGATGC TCATAGAAGT TCCTGTCCAG AGACCTGTTC TCTATCCACA
951 AGAGCAAGAA AATGAGCCAT CCATCGGGGA GTTTAATGAG CAGGTTCTGC
1001 GACTGATGCA CAGCCTTGGG ATAGATCAGC AGAAAACCAT TGAGTCTTTG
1051 CAGAACAAGA GCTATAACCA CTTTGTGCTC ATTTATTCTT TGTGGTGGA
1101 GCGCCTGAAA TCACATCGGA GCAGTTTCCC AGTGGAGCAG AGACTTGATG
1151 GCGGCCAGCG TCGGCCTAGC ACCATTGCTG AGCAAAACAG TGCCAAGGCA
1201 CAGACTGTGG GGCTCCAGT GACCATGCAT TCACCGAACA TGAGGCTGCT
1251 GCGATCTGCC CTCTCCCCC AGGCATCCAA CGTGGAGGCC TTTTCATTTT
1301 CAGCATCTGG CTGTGAGGCG GAAGCTGCAT TCATGGAAGA AGAGTGTGTG
1351 GACACTCCAA AGGTCAATGG CTGTCTGCTT GACCTGTGTC CTCTGTCTCT
1401 GGTGCGGAAG GGATGCCAGT CACTGCCCGG CAACATGATG GAGACCTCCA
1451 TTGACGAAGG GCTGGAGACA GAAGGAGAGG CCGAGGAAGA CCCCCTCAT
1501 GCCTTTGAGG CATTTAGTC CACACGAGC GGCAGAGAC GGCACACTCT
1551 GTCAGAAGTG ACCAATCAAC TGGTCTGTAT GCCTGGGGCA GGGAAAATTT
1601 TCTCCATGAA TGACAGCCCC TCCCTTGACA GTGTGGACTC TGAGTATGAT
1651 ATGGGGTCTG TTCAGAGGGA CCTGAACCTT CTGGAAGACA ACCCTTCCCT
1701 TAAGGACATC ATGTTAGCCA ATCAGCCTTC ACCCCGATG ACATCTCCCT
1751 TCATAAGCCT GAGACCTACC AACCAGCCA TGCAGGCTCT GAGCTCCAG
1801 AAACGAGAGG TCCACAACAG GTCTCCAGTG AGCTTCAGAG AGGGCCGAG
1851 AGCATCAGAT ACCTCCCTCA CCCAGGGAAT TGTAGCATT AGACAACATC
1901 TTCAGAATCT GGCTAGAACC AAAGGAATTC TAGAGTTGAA CAAAGTGCAG
1951 TTGTTGTATG AACAAATAGG ACCGGAGGCA GACCCTAACC TGGCGCCGGC
2001 GGCTCCTCAG CTCCAGGACC TTGCTAGCAG CTGCCCTCAG GAAGAAATTT
2051 CTCAGCAGCA GGAAGCGTC TCCACTCTCC CTGCCAGCGT GCATCCCCAG
2101 CTGTCCCCAC GGCAGAGCCT GGAGACCCAG TACCTGCAGC ACAGACTCCA
2151 GAAGCCACAG CTTCTGTCAA AGGCCAGAA CACCTGTGAC CTTTATTGCA
2201 AAGAACCACC GCGAGCCTT GAGCAGCAGC TGCAGGAACA TAGGCTCCAG
2251 CAGAAGCGAC TCTTTCTTCA GAAGCAGTCT CAACTGCAGG CCTATTTTAA
2301 TCAGATGCAG ATAGCAGAGA GCTCCTACCC ACAGCCAAGT CAGCAGCTGC
2351 CCTTCCCCCG CCAGGAGACT CCACCGCCTT CTCAGCAGGC CCCACGTTT
2401 AGCCTGACCC AGCCCCTGAG CCCCGTCTG GAGCCTTCCT CCGAGCAGAT
2451 GCAATACAGC CCTTTCCTCA GCCAGTACCA AGAGATGCAG CTTTCCAGCCC
2501 TGCCCTCCAC TTCCGGTCCC CGGGCTGCTC CTCTCTGCC CACGCAGCTA
2551 CAGCAGCAGC AGCCGCCACC GCCACCACC CCTCCACCAC CACGACAGCC
2601 AGGAGCTGCC CAGGCCCCCT TACAGTTCTC CTATCAGACT TGTGAGCTGC

```

```

2651 CAAGCGCTGC TTCCCCTGCG CCAGACTATC CCACTCCCTG TCAGTATCCT
2701 GTGGATGGAG CCCAGCAGAG CGACCTAACG GGGCCAGACT GTCCCAGAAG
2751 CCCAGGACTG CAAGAGGCCC CCTCCAGCTA CGACCCACTA GCCCTCTCTG
2801 AGCTACCTGG ACTCTTTGAT TGTGAAATGC TAGACGCTGT GGATCCACAA
2851 CACAACGGGT ATGTCCTGGT GAATTAGTCT CAGCACAGGA ATTGAGGTGG
2901 GTCAGGTGAA GGAAGAGTGT ATGTTCTTAT TTTTATTTCA GCCTTTTAAA
2951 TTTAAAGCTT ATTTTCTTGC CCTCTCCCTA ACGGGGAGAA ATCGAGCCAC
3001 CCAACTGGAA TCAGAGGGTC TGGCTGGGGT GGATGTTGCT TCCTCTGGT
3051 TCTGCCCCAC CACAAGTTT TCTGTGGCAA GTGCTGGAAC ATAGTTGTAG
3101 GCTGAGGCTC CTGCCCTTCG GTCGAGTGA GCAAGCTCTC GAGGGCAGCA
3151 CTGACAAATG TGTTCCTAAG AAGACATTCA GACCCAGGTC TTATGCAGGA
3201 TTACATCCGT TTATTATCAA GGGCAACCTT GGTGAAAGCA GAAAGGGTGT
3251 GTGCTATTGC ATATATATGG GGGAAAAGGC AATATATTTT TCACTGAAGC
3301 TGAGCAACCA CATATTGCTA CAAGGCAAAAT CAAGAAGACA TCAGGAAATC
3351 AGATGCACAG GAAATAAAGG AAAGCTGTGC TTTGTCTATT AATCCTAAGT
3401 TCTTAGCTGC TGATGCAAGT TGTCCCCCAA GGCCATCACA AAGCAGTGGG
3451 GCATGAGCTG TGTTCAGGG GCCACTAAAT AACAGCTGGT ACTGACCCCA
3501 GAAACCGCCT TCATCTCCAT TCGGAAGCAG GTGACACACC CCTTCAGAAG
3551 GTGCCCTGGG TTGCCGAGTG TCAGAATATA CTCAGGACTC CAGAGGTGTC
3601 ACACGTGGAA CTGACAGGAG ACCCGCCACC GTGGAGGCAG GGGGCAAGAA
3651 ACTCAAGAAC GCATCAAGAG CACCAGCCCT GGGCCAGGGA AGACAGGCTC
3701 TTCTCTGAGT TTCTCGTGGG CACTGCTGGC TTGCGGGCAG TCGGTCTCCA
3751 GGGTACCTGT TGTCTCTTTT CCGATGTAAT AACTACTTTG ACCTTACACT
3801 ATATGTTGCT AGTAGTTTAT TGAGCTTTGT ATATTTGGAC AGTTTCATAT
3851 AGGGCTTAGA GATTTTAAGG ACATGATAAA TGAACCTTTT TGTCCTATGT
3901 GAAGTGGTAG TGCGGTGCGT TTCCCCCAGA TCATGCTTTA ATCTTTCTTT
3951 TTCTGTAGAA ACCAACAGTT TCCATTATG TCAATGCTAA ATCCAAAGTC
4001 ACTTCAGAGT TTGTTTTCCTA CCATGTGGGA ATCAGCATTG TTAATTTCTG
4051 TAAAGTTTGT ACTTGTAAATG AAATGTTCAA GTATTACAGC AATATTCAAA
4101 GAAAGAACCA CAGATGTGTT AACCATTTAA GCAGATCATC TGCCAAACAT
4151 TATATTACTA ATAAACTTAA ACCAACACTT ACAATTCAGT CATCAAAGTA
4201 AGTAAAAAAT AGATGCTACA GCTAGCTAAC TGTATCCCTA GAAATGATGA
4251 ATAATTGCCC ATTTGGACAG TTAACATCCA GGTGTTACAA AGTCAGTGT
4301 AATTCTAAAG ATGATCATTT CTGCCCTTTA GAATGGCTTG TCCCATCAGC
4351 AGATGAATGT GTTAAGCACA AAGCATCTTC CTTAAAGCAC AAAGAGAGGG
4401 ACTAACTGAT GCTGCATCTA GAAAACACCT TTAAGTTGCC TTTCTCTTT
4451 GTAGTTAGCG TTCAGGCAGG TGACGTGTGG AAAGTCTAGG GGGTTCCATT
4501 CTGGCCATGC GAGCCAGCT CCTACCAACG TCGGTAACCT GAGCAGTCCC
4551 TGTGCTGGC CAGAGACTGC CTGGTCGCCA GCGCTCACCA TGGGTGCCAG
4601 GATGCTTCGC AGAGGCACTG TGCTCACGGT TGGACTGGT GTGAGTGGGA
4651 AAGGGCAGTG TGGGACTGT CATTTTGTG ATTTAATAAC ACACAGTGAA
4701 AATCCAGGAA GAATGAATTA AGCTTCTTCT GGGAGTTGTT TATTCCTGCT
4751 CGTGCTTAAG ATTGATGATT TCGTGAAATA AAGAACATCA TTTTATTAA
4801 AAAAAAAAAA AAAAAAAGG CGGCCGCTCT AGAGGATCCA AGCTTACGTA
4851 CGCGTGAAAA AAAAAAAG

```

BLAST Results

Entry HSG4921 from database EMBL:

human STS SHGC-37164.

Score = 1605, P = 1.9e-66, identities = 349/369

Entry AB018324 from database EMBL:

Homo sapiens mRNA for KIAA0781 protein, partial cds.

Score = 10725, P = 0.0e+00, identities = 2145/2145

Medline entries

No Medline entry

Peptide information for frame 1

ORF from the beginning to 2874 bp; peptide length: 959

Category: known protein

```

1 EQAERPSPKP SRAANPPARP RSCPPCLAAG PSMVMADGPR HLQRGPVRVG
51 FYDIEGTGK GNFAVVKLGR HRITKTEVAI KIIDKSQDLA VNLEKIYREV
101 QNMKMLDHPH IAKLYQVMET KSMLYLVTEY AKNGEIFDYL ANHGRLENESE
151 ARKKFWQILS AVDYCHGRKI VHRDLKAENL LLDNNMNIKI ADFGFGNFFK
201 SGELLATWCG SPPYAAPEVF EGQQYEGPQL DIWSMGVVLY VLVCGALPFD
251 GPTLPILRQR VLEGRFRIPY FMSEDCEHLI RRMLVLDPSK RLTIQIKEH

```

```

301 KMWLIEVPVQ RPYLYPQEQE NEPSIGEFNE QVLRMLHSLG IDQQTIESL
351 QNKSYNHFAA IYFLLVERLK SHRSSFPEVEQ RLDGRQRRPS TIAEQTVAKA
401 QTVGLPVTMH SPNMRLLRSA LLPQASNVEA FSFPASGCQA EAFMEEECV
451 DTPKVGNCCL DPVPPVLVRK GCQSLPSNMM ETSIDEGLT EGEAEEDPAH
501 AFEAFQSTRS GQRRHTLSEV TNQLVVMPGA GKIFSMNDSP SLDSVDSEYD
551 MGSVQRDLNF LEDNPSLKDI MLANQPSPRM TSPFISLRPT NPAMQALSSQ
601 KREVHNRSPV SFREGRRASD TSLTQGIVAF RQHLQNLART KGILELNKVQ
651 LLYEQIGPEA DPNLAPAAPQ LQDLASSCPQ EEVSQQQESV STLPASVHPQ
701 LSPRQSLETQ YLQHLQKPS LLSKAQNTCQ LYCKEPPPSL EQQLQEHRLQ
751 QKRLFLQKQS QLQAYFNQMQ IAESSYPQPS QQLPLPRQET PPSQQAAPPF
801 SLTQPLSPVL EPSSEQMQYS PFLSQYQEMQ LQPLPSTSGP RAAPPLPTQL
851 QQQQPPPPPP PPPPRQPGAA PAPLOFSYQT CELPSAASPA PDYPTPCQYP
901 VDGAAQQSDLT GPDCPRSPGL QEAPSSYDPL ALSELPGLFD CEMLDVADPQ
951 HNGYVLVN

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_15kl1, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_15kl1, frame 1

Report for DKFZphtes3_15kl1.1

```

[LENGTH]      926
[MW]           103915.77
[pI]           5.70
[OMOL]         TREMBL:AB018324_1 gene: "KIAA0781"; product: "KIAA0781 protein"; Homo sapiens
mRNA for KIAA0781 protein, partial cds. 0.0
[FUNCAT]       01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YDR477w]
8e-76
[FUNCAT]       11.01 stress response [S. cerevisiae, YDR477w] 8e-76
[FUNCAT]       30.03 organization of cytoplasm [S. cerevisiae, YDR477w] 8e-76
[FUNCAT]       98 classification not yet clear-cut [S. cerevisiae, YCL024w] 4e-58
[FUNCAT]       03.25 cytokinesis [S. cerevisiae, YDR507c] 3e-56
[FUNCAT]       03.04 budding, cell polarity and filament formation [S. cerevisiae, YDR507c]
3e-56
[FUNCAT]       30.02 organization of plasma membrane [S. cerevisiae, YDR122w] 1e-53
[FUNCAT]       03.22 cell cycle control and mitosis [S. cerevisiae, YKL101w] 3e-53
[FUNCAT]       30.10 nuclear organization [S. cerevisiae, YKL101w] 3e-53
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YPL141c] 5e-51
[FUNCAT]       03.19 recombination and dna repair [S. cerevisiae, YPL153c] 3e-42
[FUNCAT]       03.22.01 cell cycle check point proteins [S. cerevisiae, YPL153c] 3e-42
[FUNCAT]       10.99 other signal-transduction activities [S. cerevisiae, YPL153c] 3e-42
[FUNCAT]       11.04 dna repair (direct repair, base excision repair and nucleotide excision
repair) [S. cerevisiae, YPL153c] 3e-42
[FUNCAT]       03.01 cell growth [S. cerevisiae, YFR014c] 5e-42
[FUNCAT]       03.16 dna synthesis and replication [S. cerevisiae, YMR001c] 2e-34
[FUNCAT]       03.10 sporulation and germination [S. cerevisiae, YGL180w] 1e-27
[FUNCAT]       08.13 vacuolar transport [S. cerevisiae, YGL180w] 1e-27
[FUNCAT]       06.13.04 lysosomal and vacuolar degradation [S. cerevisiae, YGL180w] 1e-27
[FUNCAT]       10.02.11 key kinases [S. cerevisiae, YBL105c] 3e-26
[FUNCAT]       04.99 other transcription activities [S. cerevisiae, YER129w] 3e-26
[FUNCAT]       02.19 metabolism of energy reserves (glycogen, trehalose) [S. cerevisiae,
YPL031c] 1e-23
[FUNCAT]       01.04.04 regulation of phosphate utilization [S. cerevisiae, YPL031c]
1e-23
[FUNCAT]       04.05.01.04 transcriptional control [S. cerevisiae, YPL031c] 1e-23
[FUNCAT]       03.13 meiosis [S. cerevisiae, YOR351c] 2e-23
[FUNCAT]       10.05.11 key kinases [S. cerevisiae, YHL007c] 8e-21
[FUNCAT]       03.07 pheromone response, mating-type determination, sex-specific proteins
[S. cerevisiae, YHL007c] 8e-21
[FUNCAT]       09.01 biogenesis of cell wall [S. cerevisiae, YPL140c] 2e-20
[FUNCAT]       10.03.11 key kinases [S. cerevisiae, YLR113w] 7e-20
[FUNCAT]       04.05.01.01 general transcription activities [S. cerevisiae, YDL108w]
3e-19
[FUNCAT]       10.05.09 regulation of g-protein activity [S. cerevisiae, YBL016w] 2e-18
[FUNCAT]       10.04.11 key kinases [S. cerevisiae, YLR362w] 3e-18
[FUNCAT]       04.03.99 other trna-transcription activities [S. cerevisiae, YOR061w]
4e-18
[FUNCAT]       06.07 protein modification (glycosylation, acylation, myristylation,
palmitoylation, farnesylation and processing) [S. cerevisiae, YFL033c] 4e-17
[FUNCAT]       05.07 translational control [S. cerevisiae, YDR283c] 2e-16
[FUNCAT]       01.02.04 regulation of nitrogen and sulphur utilization [S. cerevisiae,
YNL183c] 2e-14

```

[FUNCAT] 08.99 other intracellular-transport activities [S. cerevisiae, YNL183c]
2e-14

[FUNCAT] 09.04 biogenesis of cytoskeleton [S. cerevisiae, YNL020c] 5e-14

[FUNCAT] c energy conversion [M. genitalium, MG109] 2e-12

[FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae,
YBR097w] 1e-10

[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YBR097w]
1e-10

[FUNCAT] 30.08 organization of golgi [S. cerevisiae, YBR097w] 1e-10

[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YBR097w]
1e-10

[FUNCAT] 10.04.99 other nutritional-response activities [S. cerevisiae, YJR059w]
4e-09

[FUNCAT] 01.06.10 regulation of lipid, fatty-acid and sterol biosynthesis [S.
cerevisiae, YHR079c] 1e-07

[FUNCAT] 30.07 organization of endoplasmatic reticulum [S. cerevisiae, YHR079c]
1e-07

[FUNCAT] 08.19 cellular import [S. cerevisiae, YNL154c] 2e-04

[BLOCKS] BL00415A Synapsins proteins

[BLOCKS] BL00239B Receptor tyrosine kinase class II proteins

[BLOCKS] BL00107A Protein kinases ATP-binding region proteins

[SCOP] dlgl_ 5.1.1.1.9 MAP kinase Erk2 [rat Rattus norvegicus] 3e-78

[SCOP] dlwfc_ 5.1.1.1.8 MAP kinase p38 [human (Homo sapiens)] 1e-81

[SCOP] dlkoa_2 5.1.1.1.7 (1-350) Twitchin, kinase domain [Caenorhabditis] 5e-89

[SCOP] dlkoba_ 5.1.1.1.6 Twitchin, kinase domain [california sea har] 5e-86

[SCOP] dlphk_ 5.1.1.1.5 gamma-subunit of glycogen phosphorylase kinas 3e-80

[SCOP] dlirk_ 5.1.1.2.4 insulin receptor [Human (Homo sapiens)] 6e-70

[SCOP] dlapme_ 5.1.1.1.4 cAMP-dependent PK, catalytic subunit [mouse (Mu)] 1e-95

[SCOP] dlfgka_ 5.1.1.2.3 Fibroblast growth factor receptor 1 [human (Hom)] 7e-71

[SCOP] dlydse_ 5.1.1.1.3 cAMP-dependent PK, catalytic subunit [bovine (Bo)] 2e-96

[SCOP] dlfmk_3 5.1.1.2.2 (168-437) c-src tyrosine kinase [human (Hom)] 2e-72

[SCOP] dlcdka_ 5.1.1.1.2 cAMP-dependent PK, catalytic subunit [pig (Su)] 5e-97

[SCOP] d2hckb3 5.1.1.2.1 (167-437) Haemopoietic cell kinase Hck [huma] 2e-68

[SCOP] dlcsn_ 5.1.1.1.11 Casein kinase-1, CK1 [Schizosaccharomyces pombe] 3e-53

[SCOP] dljsua_ 5.1.1.1.1 Cyclin-dependent PK [Human (Homo sapiens)] 3e-78

[SCOP] dlckia_ 5.1.1.1.10 Casein kinase-1, CK1 [rat (Rattus norvegicus)] 1e-58

[EC] 2.7.1.117 Myosin-light-chain kinase 3e-49

[EC] 2.7.1.109 [Hydroxymethylglutaryl-CoA reductase(NADPH)] kinase 4e-78

[EC] 2.7.1.38 Phosphorylase kinase 3e-41

[EC] 2.7.1.37 Protein kinase 7e-45

[EC] 2.7.1.123 Ca2+/calmodulin-dependent protein kinase 5e-42

[EC] 2.7.1.128 [Acetyl-CoA carboxylase] kinase 4e-78

[PIRKW] phosphotransferase 3e-93

[PIRKW] nucleus 2e-74

[PIRKW] calcium 2e-40

[PIRKW] transferase 3e-33

[PIRKW] duplication 2e-32

[PIRKW] tandem repeat 7e-45

[PIRKW] phorbol ester binding 4e-33

[PIRKW] zinc 4e-33

[PIRKW] ion transport 1e-32

[PIRKW] cell cycle control 1e-45

[PIRKW] serine/threonine-specific protein kinase 2e-97

[PIRKW] oncogene 1e-34

[PIRKW] phospholipid binding 2e-32

[PIRKW] autophosphorylation 2e-74

[PIRKW] brain 6e-36

[PIRKW] heterotetramer 8e-38

[PIRKW] mitosis 1e-45

[PIRKW] polymer 5e-41

[PIRKW] magnesium 6e-80

[PIRKW] ATP 2e-97

[PIRKW] polyprotein 1e-34

[PIRKW] alternative initiators 2e-31

[PIRKW] phosphoprotein 2e-74

[PIRKW] apoptosis 8e-38

[PIRKW] cGMP binding 4e-33

[PIRKW] glycoprotein 3e-36

[PIRKW] skeletal muscle 8e-38

[PIRKW] protein kinase 2e-50

[PIRKW] testis 5e-41

[PIRKW] cAMP binding 8e-38

[PIRKW] transforming protein 4e-33

[PIRKW] purine nucleotide binding 7e-52

[PIRKW] calcium binding 7e-45

[PIRKW] alternative splicing 5e-42

[PIRKW] P-loop 7e-52

[PIRKW] lipoprotein 8e-38

[PIRKW] proto-oncogene 4e-33

[PIRKW] segmentation 1e-34

[PIRKW] core protein 1e-34

[PIRKW] muscle 8e-38
 [PIRKW] myristylation 8e-38
 [PIRKW] EF hand 7e-45
 [PIRKW] cell division 3e-49
 [PIRKW] homodimer 1e-32
 [PIRKW] calmodulin binding 5e-42
 [SUPFAM] ribosomal protein S6 kinase II 1e-34
 [SUPFAM] calcium-dependent protein kinase 7e-45
 [SUPFAM] AMP-activated protein kinase 6e-80
 [SUPFAM] protein kinase akt 3e-36
 [SUPFAM] protein kinase SPK1 7e-41
 [SUPFAM] unassigned Ser/Thr or Tyr-specific protein kinases 8e-99
 [SUPFAM] Ca2+/calmodulin-dependent protein kinase 5e-42
 [SUPFAM] calmodulin repeat homology 7e-45
 [SUPFAM] cAMP receptor protein cyclic nucleotide-binding domain homology 3e-33
 [SUPFAM] protein kinase DUN1 6e-36
 [SUPFAM] protein kinase C zeta 4e-33
 [SUPFAM] Dictyostelium cAMP-dependent protein kinase catalytic chain 2e-34
 [SUPFAM] death-associated protein kinase 8e-38
 [SUPFAM] pleckstrin repeat homology 3e-36
 [SUPFAM] ankyrin repeat homology 8e-38
 [SUPFAM] protein kinase homology 8e-99
 [SUPFAM] Ca2+/calmodulin-dependent protein kinase II 6e-38
 [SUPFAM] protein kinase C zinc-binding repeat homology 4e-33
 [SUPFAM] protein kinase C delta 2e-32
 [SUPFAM] cGMP-dependent protein kinase 3e-33
 [SUPFAM] protein kinase cdrl 1e-45
 [SUPFAM] kinase-related transforming protein 2e-50
 [SUPFAM] Ca2+/calmodulin-dependent protein kinase I 8e-42
 [SUPFAM] kinase interaction domain homology 7e-41
 [SUPFAM] gag-akt polyprotein 1e-34
 [PROSITE] PROTEIN_KINASE_ATP 1
 [PROSITE] MYRISTYL 3
 [PROSITE] AMIDATION 2
 [PROSITE] CAMP_PHOSPHO_SITE 4
 [PROSITE] CK2_PHOSPHO_SITE 15
 [PROSITE] TYR_PHOSPHO_SITE 2
 [PROSITE] PKC_PHOSPHO_SITE 10
 [PROSITE] ASN_GLYCOSYLATION 2
 [PROSITE] PROTEIN_KINASE_ST 1
 [PFAM] Eukaryotic protein kinase domain
 [KW] Irregular
 [KW] 3D
 [KW] LOW_COMPLEXITY 12.31 %

SEQ MVMADGPRHLQRGPVRVGFYDIEGLGKGNFAVVKLGRHRITKTEVAIKIIDKSQLDVAVN
 SEG
 1ctpEEEECTTTEEEEEETTTTEEEEEEEHHHHHHHC

SEQ LEKIYREVQIMKMLDHPHIKLYQVMETKSMYLYVTEYAKNGEIFDYLANHGRLNESEAR
 SEG
 1ctpE HHHHHHHHHHHHCCCTTTBCCEEEEEETTEEEEEECTTTTBHHHHHHHHHCCCHHHHH

SEQ RKFWQILSAVDYCHGRKIVHRDLKAENLLDNNMNIKIADFGFGNFFKSGELLATWCGSP
 SEG
 1ctpE HHHHHHHHHHHHHHCCCECCCGGGEETTTTCEEECTTTTEETT-TTBC-CCCCCG

SEQ PYAAPEVFEGQQYEGPQLDIWSMGVVLYVLVCGALPFDGPTLPILRQRVLEGRFRIPIYFM
 SEG
 1ctpE GGCCHHHHHCCBCB-HHHHHHHHHHHHHHHHCCCTTTTTHHHHHHHHHHCCCCCTTTT

SEQ SEDCEHLIRRLVLDPSPKRLTIAQIKEHKWMLIEVPVQRPVLYPQEQENEPSIGEFNEQV
 SEG
 1ctpE CHHHHHHHHHHTTTTGGGTTTHHHHHHCGG.....

SEQ LRLMHSLSGIDQQKTIESLQNKSYNHFAAIYFLLVERLKSHRSSFPVEQRLDGRQRRPSTI
 SEG
 1ctpE

SEQ AEQTVAKAQTVGLPVTMHSNMLLRALLPQASNVEAFSFPASGCQAEAAFMEEECVDT
 SEG
 1ctpE

SEQ PKVNGCLDPPVPLVVRKGCQSLPSNMETSIDEGLETEGEAEEDPAHAFAEFQSTRSGQ
 SEGxxxxxxx
 1ctpE

SEQ RRHTLSEVTNQLVVMGAGKIFSMNDSPSLDSVDSEYDMGSVQRDLNLFLEDNPSLKDIML
 SEG
 1ctpE

```

SEQ  ANQPSPRMTSPFISLRPTNPAMQALSSQKREVHNRSVPSFREGRASDTSLTQGIVAFRQ
SEG  .....
lctpe .....

SEQ  HLQNLARTKGILELNKVQLLYEQIGPEADPNLAPAPQLQDLASSCPQEEVSQQQESVST
SEG  .....XXXXXXXXXXXXXXXXXXXXX.....XXXXXXXXXXXXX
lctpe .....

SEQ  LPASVHPQLSPROSLETQYLQHLRLQKPSLLSKAQNTCOLYCKEPPRSLEQQLQEHRLQOK
SEG  .....XXXXXXXXXXXXX
lctpe .....

SEQ  RLFLQKQSQLQAYFNQMQUIAESSYPQPSQQLPLPRQETPPPSQQAPPFSLTQPLSPVLEP
SEG  .....XXXXXXXXXXXXXXXXXXXXX.....
lctpe .....

SEQ  SSEQMQYSPFLSQYQEMQLQLPSTSGPRAAPPLPTQLQQQPPPPPPPPPPRQPGAAPA
SEG  .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
lctpe .....

SEQ  PLQFSYQTCELPSAASPADYPTPCQYPVDGAQQSDLTGPDCPRSPGLQEAPSSYDPLAL
SEG  xxx.....
lctpe .....

SEQ  SELPGLFDCEMLDAVDPQHNGYVLVN
SEG  .....
lctpe .....

```

Prosites for DKFZphtes3_15k11.1

PS00001	115->119	ASN_GLYCOSYLATION	PDOC00001
PS00001	320->324	ASN_GLYCOSYLATION	PDOC00001
PS00004	258->262	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	355->359	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	481->485	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	584->588	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	257->260	PKC_PHOSPHO_SITE	PDOC00005
PS00005	339->342	PKC_PHOSPHO_SITE	PDOC00005
PS00005	420->423	PKC_PHOSPHO_SITE	PDOC00005
PS00005	475->478	PKC_PHOSPHO_SITE	PDOC00005
PS00005	534->537	PKC_PHOSPHO_SITE	PDOC00005
PS00005	545->548	PKC_PHOSPHO_SITE	PDOC00005
PS00005	554->557	PKC_PHOSPHO_SITE	PDOC00005
PS00005	567->570	PKC_PHOSPHO_SITE	PDOC00005
PS00005	579->582	PKC_PHOSPHO_SITE	PDOC00005
PS00005	670->673	PKC_PHOSPHO_SITE	PDOC00005
PS00006	42->46	CK2_PHOSPHO_SITE	PDOC00006
PS00006	54->58	CK2_PHOSPHO_SITE	PDOC00006
PS00006	128->132	CK2_PHOSPHO_SITE	PDOC00006
PS00006	292->296	CK2_PHOSPHO_SITE	PDOC00006
PS00006	359->363	CK2_PHOSPHO_SITE	PDOC00006
PS00006	394->398	CK2_PHOSPHO_SITE	PDOC00006
PS00006	450->454	CK2_PHOSPHO_SITE	PDOC00006
PS00006	458->462	CK2_PHOSPHO_SITE	PDOC00006
PS00006	484->488	CK2_PHOSPHO_SITE	PDOC00006
PS00006	503->507	CK2_PHOSPHO_SITE	PDOC00006
PS00006	515->519	CK2_PHOSPHO_SITE	PDOC00006
PS00006	534->538	CK2_PHOSPHO_SITE	PDOC00006
PS00006	579->583	CK2_PHOSPHO_SITE	PDOC00006
PS00006	878->882	CK2_PHOSPHO_SITE	PDOC00006
PS00006	893->897	CK2_PHOSPHO_SITE	PDOC00006
PS00007	672->680	TYR_PHOSPHO_SITE	PDOC00007
PS00007	100->108	TYR_PHOSPHO_SITE	PDOC00007
PS00008	372->378	MYRISTYL	PDOC00008
PS00008	871->877	MYRISTYL	PDOC00008
PS00008	905->911	MYRISTYL	PDOC00008
PS00009	134->138	AMIDATION	PDOC00009
PS00009	582->586	AMIDATION	PDOC00009
PS00107	26->50	PROTEIN_KINASE_ATP	PDOC00100
PS00108	138->151	PROTEIN_KINASE_ST	PDOC00100

Pfam for DKFZphtes3_15k11.1

HMM_NAME Eukaryotic protein kinase domain

HMM		*YeigRiIGeGsFGtVYkCiWr.TGeIvAIKIIkkrsms.....FlREI	
		Y I++++G+G+F++V++++R T +VAIKII+K++++ + RE+	
Query	20	YDIEGTLGKGNFAVVKLGRHRITKTEVAIKIIDKSQLDAPVNLEKIYREV	68
HMM		qIMRrLnHPNIIRFYDwFedddDHIYMIMEYMeGGDLFDYIrrngpMsEw	
		QIM++L+HP+II++Y ++E +++ +Y+++EY+ +G++FDY+ ++G+++E	
Query	69	QIMKMLDHPHIKLYQVME-TKSMlyLVTEYAKNGEIFDYLANHGRLNES	117
HMM		eIrfIMyQILrGMeYLHSMgIIHRDLKPENILIDeNgqIKicDFGLARqM	
		E+R+ ++QIL++++Y+H ++I+HRDLK+EN+L+D+N++IKI+DFG+ ++	
Query	118	EARRKFWQILSAVDYCHGRKIVHRDLKAENLLDNMMNIKIADFGFGNFF	167
HMM		nnYerMttfCGTPWYMAPEVIimg.nyYttkVDMWSFGCILWEMMTGep	
		+++E++ T CG+P+Y APEV +G +Y +++ D+WS+G++L+ +++G +	
Query	168	KSGELLATWCGSPPYA-APEV-FEGQQYEGPQLDIWSMGVLYVLVCGAL	215
HMM		PFyddnMemImrIiqrfrfwpnCSeElyDFMrwCWnyDPekRPTFrQI	
		PF++ ++ + + +++ R+++++ +SE++ +++R+++ +DP+KR+T+ QI	
Query	216	PFDGPTLPILRQRVLEGRFRIPYFMSDCEHLIRRMLVLDPSKRLTIAQI	265
HMM		LnHPWF*	
		+H W+	
Query	266	KEHKWM	271

DKFZphtes3_17f10

group: testes derived

DKFZphtes3_15j18 encodes a novel 710 amino acid protein with weak similarity to neurofilament proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to neurofilament proteins

Sequenced by GBF

Locus: unknown

Insert length: 2533 bp

Poly A stretch at pos. 2507, no polyadenylation signal found

```
1  CTTCA GTTCA ACTAAAAATG GACAGATCTC AGCAGACCAG CCGTACAGGA
51 TACTGGACCA TGATGAACAT CCCCCTGTGA GAAAAAGTGG ACAAGGAACA
101 ACAGACATAC TTTAGTGAAT CAGAAATAGT GGTATTTCCT AGGCCAGATA
151 GTTCTTCTAC AAAGTCAAAG GAAGATGCCC TGAACATAAA ATCGTCGGGA
201 AAGATTTTTG CTAGTGAACA CCCTGAATTT CAACCAGCAA CAAACAGCAA
251 TGAAGAAATT GGGCAGAAAA ATATCAGCAG AACTTCATTT ACTCAGGAGA
301 CTAAGAAAGG TCCCCCAGTA CTTTATAGAAG ATGAGCTTAG GGAAGAAGTA
351 ACTGTACCTG TTGTACAAGA AGGTTCTGCT GTTAAAAAAG TGGCTTCTGC
401 TGAATAGAGC CCTCCATCAA CAGAAAAATT CCCAGCTAAA ATACAGCCTC
451 CATTAGTTGA AGAGGCCACT GCTAAAGCGG AGCCCAGACC TGCTGAAGAG
501 ACCCATGTCC AAGTACAGCC ATCAACTGAA GAGACTCCTG ATGCTGAGGC
551 AGCCACTGCA GTTGC GGAGA ATTCTGTTAA AGTTCAGCCT CCACCTGCTG
601 AAGAGGCCCC TTTAGTGGAG TTCTCTGCTG AAATTCAGCC TCCATCAGCT
651 GAAGAGTCTC CTTCTGTAGA GCTTCTGGCT GAAATTCCTG CTCCATCAGC
701 TGAAGAGTCC CCTTCAGAAG AGCCTCCTGC TGAATTTCTG CCTCCACCAG
751 CTGAAAAGTC TCCTTCAGTA GAGCTTCTTG GTGAAATTCG GTCTCCCTCA
801 GCACAAAAGG CTCCCATTGA AGTACAGCCT TTACCAGCTG AGGGCGCCCT
851 TGAAGAGGCC CCAGCTAAAG TAGAGCCTCC CACTGTTGAA GAGACCCTTG
901 CTGAAGTTCA GCCTCTATTA CCTGAAGAGG CTCCTAGAGA AGAGGCTCGA
951 GAACCTCAGC TTTCACACGC TATGGAGACC CCTGCAGAAG AGGCTCCTAC
1001 TGAATTTTCA GCTCCATTAC CTAAGAGAGC CACTGCAGAA GAGGCCCTCTG
1051 CTGAAATTTA GCTTCTAGCA GCTACGGAGC CTCCTGCAGA TGAAACTCCT
1101 GCCGAAGCTC GGTCTCCACT ATCTGAGGAG ACTTCTGCAG AAGAGGCTCA
1151 TGCTGAAGTT CAATCTCCAT TAGCTGAAGA GACCACTGCA GAAGAGGCCT
1201 CTGCTGAAAT TCAGCTTCTA GCAGCTATAG AGGCTCCTGC AGATGAAACT
1251 CTTGCTGAAG CTCAGTCTCC ACTATCTGAG GAGACTTCTG CAGAAGAGGC
1301 TCCTGCTGAA GTTCAGTCTC CATCAGCTAA GGGAGTTTCT ATAGAAGAGG
1351 CCCCTCTTGA GCTTCAGCCT CCATCAGGTG AAGAGACCAC TGCAGAAGAG
1401 GGCTCTGCTG CAATTCAGCT TCTAGCAGCT ACAGAGGCTT CTGCAGAAGA
1451 GGCTCTGCTG GAAGTTCAGC CTCCACCAGC TGAGGAGGCC CCCGCTGAAG
1501 TTCAGCCTCC ACCAGCTGAG GAGGCCCCCG CTGAAGTTCA GCCTCCACCA
1551 GCTGAGGAGG CCCCCTGCTA AGTTCAGCCT CCACCAGCTG AGGAGGCCCC
1601 CGCTGAAGTT CAGCCTCCAC CAGCTGAGGA GGCCCCCGCT GAAGTTCAGC
1651 CTCCACCAGC TGAGGAGGCC CCCTCTGAAG TTCAGCCTCC ACCAGCTGAG
1701 GAGGCCCTCTG CTGAAGTTCA GTCTCTACCA GCTGAGGAGA CTCCCTATAGA
1751 AGAGACCCCTT GCTGCAGTAC ACTCTCCCCC AGCTGATGAT GTCCCTGCAG
1801 AAGAGGCCCTC CGTTGACAAA CATTCCCCAC CAGCTGATTT GCTTCTGACT
1851 GAGGAGTTTC CTATAGGAGA GGCCTCTGCT GAAGTTTCAC CTCCACCATC
1901 TGAACAAACC CCTGAAGATG AGGCTCTGGT AGAGAATGTG TCTACAGAAT
1951 TTCAGTCAAC GCAGGTGGCA GGAATTCCAG CAGTAAAAAT AGGATCGGTT
2001 GTTTTGGAAG GTGAAGCAAA ATTTGAAGAG GTTTCAAAAA TCAATTCTGT
2051 CCTTAAAGAT TTGTCTAATA CCAATGATGG ACAGGCTCCC ACTCTTGAAA
2101 TAGAAAGTGT TTTTCATATA GAATTAATAA AACGTCCTCC TGAACGTAG
2151 TCAGGTTGTA CCTAAGCTAG CAATCAGAAG CTACATGGTT TTGGAAGAAC
2201 ATACTTTTGA AAAGGGTGGG CAGCAGGAAG TAGCTTTGTC AATAAGGCAA
2251 ATTAAGGGGG ACCCAAGAC TTGGAATACA GGTGGAATAA TGAACAATAA
2301 AAACGTGAGC AGCATAAAAT TACTTGTGTT AATTTCATTC AAATTTATGG
2351 CATGAAAAAT ACCTATTTTG AAAGTAAGTT TATAATTGAA AAAAAATTGCT
2401 TAAAAATATC TTCTTACAGT AAACCTGTTG ACACGAGTAA AGTTTAAATCT
2451 GCAGCCATCT TTTCTGTCTT TTGCCTTCCC TTTATAAGTA AATATAGTTT
2501 CTAGTGAAAA AAAAAAAAAA AAAAAAAAAA AAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 18 bp to 2147 bp; peptide length: 710
Category: similarity to known protein
Classification: unclassified

```

1 MDRSQTSRT GYWTMMNIPP VEKVDKEQQT YFSESEIVVI SRPDSSTKS
51 KEDALKHKSS GKIFASEHPE FQPATNSNEE IGQKNISRTS FTQETKKGPP
101 VLLEDELREE VTPVVQEGS AVKKVASAEI EPPSTKFPFA KIQPPPLVEEA
151 TAKAEPRPAE ETHVQVPST EETPDAAEAT AVAENSVKVQ PPPAEAPLV
201 EFPAEIQPPS AEESPSVELL AEILPPSAEE SPSEEPFAEI LPPPAEKSPS
251 VELLGEIRSP SAQKAPIEVQ PLPAEGALEE APAKVEPPTV EETLAEVQPL
301 LPEEAPREEA RELQLSTAME TPAAEAPTEF QSPLPKETTA EEASAEIQLL
351 AATEPPADET PAEARSPLSE ETSAAEAHAE VQSPLAETT AEEASAEIQL
401 LAAIEAPADE TPAAEQSPLS EETSAAEAPA EVQSPSAKGV SIEEAPLELQ
451 PPSGEETTAE EASAAIQLLA ATEASAEAP AEVQPPPAEE APAEVQPPPA
501 EEAPAEVQPP PAEEAPAEVQ PPPAEAPAE VQPPPAEAP AEVQPPPAEE
551 APSEVQPPPA EEAPAEVQSL PAETPIEET LAAVHSPPAD DVPAAEASVD
601 KHSPPADLLL TEEFPIGEAS AEVSPPPSEQ TPEDEALVEN VSTEFQSPQV
651 AGIPAVKLGS VVLEGEAKFE EVSKINSVLK DLSNTNDGQA PTLIEISVFH
701 IELKQRPPEL

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_17f10, frame 3

PIR:A37221 neurofilament triplet H protein - rat, N = 1, Score = 480, P = 7.4e-43

TREMBL:RNNFLH_1 Rat heavy neurofilament subunit (NF-H) mRNA, 3' end., N = 1, Score = 475, P = 1e-42

>PIR:A37221 neurofilament triplet H protein - rat
Length = 1,072

HSPs:

Score = 480 (72.0 bits), Expect = 7.4e-43, P = 7.4e-43
Identities = 185/622 (29%), Positives = 320/622 (51%)

```

Query:   33 SESEIVVISRPDSSTKSKEDALKHKSSGKIFASEHPEFQPATNSNEEIGQKNISRTSFT 92
          SE +I V+ + + + +E + + + + + E E Q E G + + TS
Sbjct:  436 SEEKIKVVEKSEKETVIVVEEQTEEIQTVEEVTEEDKEAQGEEEEAEEGGEEAATTSP 495

Query:   93 QETKKGPPVLLEDELREEVTVPVVQEGSAVKKVASAEIEPPSTKFPFAKIQPPPLVEEATA 152
          E P + ++EE P + A K + AE + P+ K PA+++ P ++ A
Sbjct:  496 AEEAASPEKETKSPVKEEAKSPAEAKSPAEAK-SPA EAKSPAEVKS PAEVKSPAEAKSPA 554

Query:  153 KAEPRPAETHVQVPSTETPDAAEATAVAENSVKVQPPPAEAP-LVEFPAEIQPPSA 211
          +A+ PAE V+ P+T ++P + A A++ +V+ P ++P + PAE + P+
Sbjct:  555 EAKS-PAE---VK-SPATVKSPA EAKSPA EAKSPA EVKSPATVKSPGEAKSPA EAKSPA 609

Query:  212 EESP-SVELLAEILPPSAEESPSE-EPPAEILPPPAEKSPS-VELLGEIRSPSAQKAPIE 268
          +SP + AE P++ +SP E + PAE P KSP+ V+ E +SP+ K+P+
Sbjct:  610 VKSPVEAKSPA EAKSPASVKSPGEAKSPA EAKSPA EVKSPATVKSPVEAKSPA EVKSPVT 669

Query:  269 VQPLPAEGALEEAPAKVEPPTVEETLAEVQPLLP EAPREEARELQLSTAMETPAE-EAP 327
          V+ PAE ++P +V+ P ++ +E + ++P E A+ ++PAE ++P
Sbjct:  670 VKS-PAEA---KSPVEVKS PASVKSP EAKSPA EAKSPA EVKSPATVKSPVEAKSPA EAKSPA 721

Query:  328 TEFQSP LPKETTAEASAEIQLLAATEPPAD-ETPAEARSPLSEETSAAEAHA EVQS--- 383
          E + P ++ AE S A + PA+ ++PAEA+SP+ E S E+A + V+
Sbjct:  722 AEAKPPAEAKSPA EAKSP-----AEAKSPA EAKSPA EAKSPV-EVKSP EAKSPA EAKSPA 775

Query:  384 PLAEETTAEASAEIQLLAIEAPAD-ETPAEASPLSEET-SAE EAPA-EVQSPSAKGV 440

```

Sbjct: 776 LAE + E+A + ++ I+ PA+ ++P +A+SP+ EE S E+A +V+SP AK
 SLAEAKSPEKAKSPVK--EEIKPPAEVKSPEKAKSPMKEEAKSPEKAKTLDVKSPEAKTP 833

Query: 441 SIEEA--PLELQPPSGEETTA-EEASAAIQLLAATEASA---EEAPAEVQPPPAEEAPAE 494
 + EEA P +++ P ++ A EEA + + TE A EE + V+ A+E P +

Sbjct: 834 AKEEAKRPADIRSEQVKSPEKEAKSPEKEETRTKVPKKEEVKSPVEEVKAKEPPKK 893

Query: 495 VQPPPAEEAP-AEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPS 553
 V+ P EV+ +EAP E Q P AEE + P +++P E + EEA

Sbjct: 894 VEEKTPATPKTEVKESKKDEAPKEAQKPKAEKEPLTEKP--KDSGGEAKK---EEAKE 948

Query: 554 EVQPPPAEEAPAEV---QSLP---AEETPIEETL--AAVHSPADDVPAAEASVD-KHS 603
 + P EE PA++ ++ P AE+ +E + P ++VPA D K

Sbjct: 949 KKAAPPEETPAKLGVEAKPKKAEDAKAKEPSKPKSEKEKPKKEVPAAPEKKDTKEE 1008

Query: 604 PPADLLLTTEFFIGEASAEVSPP--PSEQT-PEDEALVENVSTEFQSPQ 649
 + EE P +A A+ P E + P+ E ++ ST+ + Q

Sbjct: 1009 KTTESKKPEKPKMQAKAKEEDKGLPQEPSPKPKTEKAEKSSSTDQKDSQ 1057

Score = 473 (71.0 bits), Expect = 4.8e-42, P = 4.8e-42
 Identities = 184/628 (29%), Positives = 310/628 (49%)

Query: 18 IPPVEKVDKEQQTYSSEIIVVISRP---DSSSTKSKEDALKHKSSGKIFASEHPFQPA 74
 I VEK +KE ++E + ++ + E+ + + G+ A+ P + A

Sbjct: 440 IKVVERSEKETVIVEEQTEEQVTEEVTEEDKEAQGEEEEAEEGGEEAATTSPPAEEA 499

Query: 75 TNSNEEIQGNISRTSFTQETKKGPPVLEDELREEVTVPVVQEGSAVKKVASAEIEPPS 134
 + +E + + + + K P E + E P + A K + AE + P+

Sbjct: 500 ASPEKET-KSPVKEEAKSPAEEKSPA---EAKSPAEEKSPAEVKSPA EVK-SPAEEKSPA 554

Query: 135 TEKFPKAIQPPPLVEEATAKAEPRPAEETHVQVQ-PSTEETPDAAATAVAENSVKVQPPP 193
 K PA+++ P ++ A+A+ ++ +V+ P+T +P+ + A A++ +V+ P

Sbjct: 555 EAKSPA EVKSPATVKS PAEAKSPA EVKSPATVKS PGEAKSPA EVKSPA EVKSPV 614

Query: 194 AEEAPL-VEFP AEIQQPSAEESPS-VELLAELPPSAEESPS-EPPAEILPPPAEKSPS 250
 ++P + PA ++ P +SP+ + AE+ P+ +SP E + PAE+ P KSP+

Sbjct: 615 EAKSPA EVKSPASVKS PGEAKSPA EVKSPATVKS PVEAKSPA EVKSPVTVKSPA 674

Query: 251 -VELLGEIRSPSAQKAPIEVQ-PLPAEGALE-EAPAKVEPPTVEETLAEVQPLLPEEAPR 307
 + E++SP++ K+P E + P A+ E ++P + P ++ AE +P ++P

Sbjct: 675 EAKSPVEVKS PASVKS PSEAKSPAGAKSPA EAKSPV VAKSPA EAKSPA EAKPPAEAKSPA 734

Query: 308 EEARELQLSTAME--TPAE-EAPTEFQSP---LP-KE---TTAEASAEIQLLAATE-- 354
 E + + E +PAE ++P E +SP P KE + AE S E E

Sbjct: 735 EAKSPA EVKSPA EVKSPA EVKSPA EVKSPA EVKSPA EVKSPA EVKSPA EVKSPA 794

Query: 355 -PPAD-ETPAEARSPLSEET-SAE EAHAEVQSPLAEETTAEAS--AEIQLLAIEAPA 408
 PPA+ ++P +A+SP+ EE S E+A +V+SP A+ EEA A+I+ +++PA

Sbjct: 795 KPPEVKSPEKAKSPMKEEAKSPEKAKTLDVKSPEAKTPAKEEAKRPADIRSEQVKSPEA 854

Query: 409 DETPAEQAQSPLEETSAAE-APA--EVQSPSAKGVSI EEA PLQPPSGEETTAEASAA 465
 E EA+SP EET E+ AP EV+SP +EE + +PP E EE + A

Sbjct: 855 KE---EAKSPEKEETRTKVPKKEEVKSP-----VEEVKAK-EPPKKVE---EEKTPA 901

Query: 466 IQLLAATEASAEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAE 525
 E+ +EAP E Q P AEE + P +++P E + A+E A P E

Sbjct: 902 TPKEVKESKKDEAPKEAQKPKAEKEPLTEKP--KDSGGEAKKEEAKKAAA---PEE 956

Query: 526 EAPAEV---QPPPAEEAPAEVQPPPAEEAPSEVQPPPAEEAPAEVQSLPAEETPIEETL 581
 E PA++ + P E+A P++ PSE + P EE PA + +E E+

Sbjct: 957 ETPAKLGVEEAKPKKAEDAKAKEPSK--PSEKEKPKKEVPAAPEKKDTKEETTESK 1014

Query: 582 AAVHSPADDVPAAEASVDKHSPPADLL-LTEFFIGEASAEVSPPPSEQTPEDEA 636
 P EE DK P TE+ ++ + PSE+ PED+A

Sbjct: 1015 KPPEKPKMQAKAKEE---DKGLPQEPSPKPKTEKAEKSSSTDQKDSQSPSEKAPEDKA 1067

Score = 421 (63.2 bits), Expect = 3.7e-36, P = 3.7e-36
 Identities = 162/540 (30%), Positives = 275/540 (50%)

Query: 135 TEKFPKAIQPPPLVEEATAKAEPR-----PAEETHVQVQSTEETPDAAATAVAENSVKV 189
 TE P KI P + K+E + +E+ V V+ TEE E T E +

Sbjct: 419 TEGLP-KI-PSMSTHIKVKSEKIKVVEKSEKETVIVEEQTEEQVTEEVTE--EEDKEA 474

Query: 190 QPPPAEEAPLVEFP AEIQQPSAEESPSVELLAELPPSAEE--SPSE-EPPAEILPPPAE 246
 Q EEA A P AEE+ S E E P EE SP+E + PAE P

Sbjct: 475 QGEEEEAEEGGEEAATTSPPAEEAASPE--KETKSPVKEEAKSPA EAKSPA EAKSPA EAKSPA 532

Query: 247 KSPSVELLGEIRSPSAQKAPIEVQPLPAEGALEEAPAKVEPPTVEETLAEVQPLLPEEAP 306
 KSP+ E++SP+ K+P E + PAE ++PA+V+ P ++ AE + ++P

Sbjct: 533 KSPA-----EVKSPA EVKSPA EAKS-PAEA---KSPA EVKSPATVKS PAEAKSPA EAKSP 583

Query: 307 REEARELQLSTAME--TPAE-EAPTEFQSPLPKETTAEASAEIQLLAATEPPAD-ETP 361
 E + + E +PAE ++P E +SP+ ++ AE S A ++ + PA+ ++P
 Sbjct: 584 AEVKSPATVKSPGEAKSPAEEKSPAENVKSPVEAKSPAEEKSPASVKSPGEAKSPAEEKSP 643

Query: 362 AEARSPLSEETSAE-EAHAEVQSPLAEETTAEASAEIQLLAIEAPAD-ETPAEQSPL 419
 AE +SP + ++ E ++ AEV+SP+ ++ AE A + ++ +++PA ++P+EA+SP
 Sbjct: 644 AEVKSPATVKSPVEAKSPAENVKSPVTVKSPA-EAKSPVE----VKSPASVKSPSEAKSP- 697

Query: 420 SEETSAEEAPAEVQSPS-AGVSIIEAPLELQPPSGEETTAEASAAIQLLAATEASAE 478
 + ++PAE +SP AK + ++P E +PP+ ++ AE S A A + A A+
 Sbjct: 698 ----AGAKSPAEEKSPVAKSPAEEKSPAEEKPPAEAKSPAEEKSPA-E-AKSPA-EAK- 749

Query: 479 APAEVQPPPAEEAPAEVQPPPAEEAP--AEVQPPPAEEAP--EVQPPPAEEAPAEVQPP 534
 +PAE + P ++P + + P E A AE + P ++P E++PP ++P + + P
 Sbjct: 750 SPAEAKSPVEVKSPKAKSPVKEGAKSLAEAKSPKAKSPVKEEIKPPAEVKSPEKAKSP 809

Query: 535 PAEEAPAEVQPPPAEEAPSEVQPPPAEEA--PAEVQSLPAEETPIETLAHVHSPADDV 592
 EEA + + + E + P EEA PA+++S ++P +E SP ++
 Sbjct: 810 MKEEAKSPEKAKTLDVKSPEAKTPAKEEAKRPADIRSPEQVKSPAEE--AKSPEKEET 866

Query: 593 PAEEASVDKHS--PPADLLLTEEFPIGEASAEVSPPPSEQTPEDALVENVSTEFQSPQV 650
 E+ + K P + + +E P + E P + +T E+ + E Q P+
 Sbjct: 867 RTKKVAPKKEEVKSPVEEVKAKEPP--KKVEEKTATPKTEVKESKKDEAPKEAQKPKA 924

Query: 651 AGIPAVKLGSVVLEGEAKFEEVSK 674
 + GEAK EE +
 Sbjct: 925 EEKEPLTEKPKDSPGEAKKEEAK 948

Score = 406 (60.9 bits), Expect = 1.7e-34, P = 1.7e-34
 Identities = 123/390 (31%), Positives = 213/390 (54%)

Query: 308 EEARELQLSTAMETPAEEAPTEFQSPLPKETTAEASAEIQLLAATEPPADETPA---EA 364
 E+ E+Q++ E EE E Q +E AEE E A T PPA+E + E
 Sbjct: 455 EQTEEIQVT---EEVTEEDKEAQGE--EEEEAEEGGEEA---ATTSPPAEEAASPEKET 506

Query: 365 RSPLSEETSAEEAHAEVQSPLAEETTAEASAEIQLLAIEAPAD-ETPAEQSPLSEE 422
 +SP+ EE + AE +SP ++ AE S AE++ A +++PA+ ++PAEA+SP +
 Sbjct: 507 KSPVKEEAKSP---AEAKSPAEEKSPAEEKSPAENVKSPAENVKSPAEEKSPAEEKSPAENVK 563

Query: 423 TSAE-EAPAEVQSPS-AGVSIIEAPLELQPPSGEETTAEASAAIQLLAATEASAEAP 480
 + A ++PAE +SP+ AK + ++P ++ P GE + EA + ++ + EA ++P
 Sbjct: 564 SPATVKSPAEEKSPAEEKSPAENVKSPATVKSP-GEAKSPAEEKSPAENVKSPVEA---KSP 619

Query: 481 AEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAP 540
 AE + P + ++P E + P ++PAEV+ P ++P E + P ++P V+ P ++P
 Sbjct: 620 AEAKSPASVKSPGEAKSPAEEKSPAENVKSPATVKSPVEAKSPAENVKSPVTVKSPAEEKSP 679

Query: 541 AEVQPPPAEEAPSEVQPPPAEEAPAEVQSLPAEETPIETLAHVHSPPAD-DVPAEEASV 599
 EV+ P + ++PSE + P ++PAE +S ++P E A PPA+ PAE S
 Sbjct: 680 VEVKSPASVKSPSEAKSPAGAKSPAEEKSPVAKSPAEEKSPAEEKPPAEAKSPAEEKSP 739

Query: 600 DKHSPPADLLLTEEFPIGEASAEVSPPPSEQTPEDALVENVSTEFQSPQVAGIPAVKLG 659
 + PA+ E ++ EV P ++P E +++ E +SP+ A P VK
 Sbjct: 740 AEAKSPAEEKSPA-E-AKSPVEVKSPKAKSPVKEG-AKSLA-EAKSPEKAKSP-VK-E 792

Query: 660 SVVLEGEAKFEEVSKINSVLKDLNNTNDGQAPTLEIES 697
 + E K E +K S +K+ + + + A TL+++S
 Sbjct: 793 EIKPPAEVKSPEKAK--SPMKEEAKSPE-KAKTLDVKS 827

Score = 255 (38.3 bits), Expect = 5.5e-18, P = 5.5e-18
 Identities = 124/420 (29%), Positives = 199/420 (47%)

Query: 252 ELLGEIRSPSAQKAPIEVQPLPA-----EGALEEAPAKVEPPTVEETLAEVQPLLPEEAP 306
 ELLG+I+ A +A + + A AL E A++E TV+ TL +
 Sbjct: 236 ELLGQIQCGAAQQAQAEARDALKCDVTSALREIRAQLEGHTVQSTLQSEEWFRVRLDR 295

Query: 307 REEARELQLSTAMETPAEEAPTEFQSPLPKETTAEASAEIQLLAATEPPADETPAEARS 366
 EA ++ + AM + EE TE++ L TT E++ L +T+ + +E
 Sbjct: 296 LSEAAKVN-TDAMRSAQEIEI-TEYRRQLQARTT-----ELEALKSTKESLERQRSELED 347

Query: 367 PLSEE-TSAEEAHAEVQSPLAEETTAEASAEIQLLAIEAPAD-ETPAEQSPLSEE 422
 + S +A ++ + L T E A+ E Q L ++ D E A + EE
 Sbjct: 348 RHQVDMASYQDAIQQLDNEL-RNTKWEMAAQLREYQDLLNVKMLDIEIAAYRKLLGEE 406

Query: 423 TSAEEAPAEV-----QSPS-AGVSIIE-EAPLELQPPSGEETT-AEEASAAIQLLA-A 471
 P+ + PS + + ++ E +++ S +ET EE + IQ+
 Sbjct: 407 CRIGFGPSPFSLTEGLPKIPSMSTHIKVKSEEKIKVVEKSEKETVIVEEQTEEIQTVEEV 466

Query: 472 TEASAEAPAEVQPPPAEEAPAEVQPPPAEEAP--PPAEAPAEVQPPPAEEA--PAEVQPPPA 524
 TE +EA E + AEE E PPAEEA + E + P EEA PAE + P
 Sbjct: 467 TEEDKEAQGE-EEEEAEEGGEEAATTSPPAEEAASPEKETKSPVKEEAKSPAEEKSPA-E 525

Query: 525 EEAPAEVQPPPAEEAPAEVQPPPAEEAPSEVQPPPAEEAPAEVQSLPAE-ETPIE-ETLA 582
 ++PAE + P ++PAEV+ P ++P+E + P ++PA V+S PAE ++P E ++ A
 Sbjct: 526 AKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPA 584

Query: 583 AVHSPADDVPAEEASVDKHSPPADLLLEEFPIGEASAEVSPFPSEQTP-EDEALVENV 641
 V SP P E S + PA++ E ++ AE P S ++P E ++ E
 Sbjct: 585 EVKSPATVKSPGEAKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPA 641

Query: 642 S-TEFQSPQVAGIP 654
 S E +SP P
 Sbjct: 642 SPAEVKSPATVKSP 655

Score = 253 (38.0 bits), Expect = 9.0e-18, P = 9.0e-18
 Identities = 115/364 (31%), Positives = 166/364 (45%)

Query: 110 EVTVPVQEGSAVKKVASAEIEPPSTEFPAKIQPPLVEEATAKAEPRAE-ETHVQVQ- 167
 E PVV + A K + AE +PP+ K PA+ + P ++ A+A+ PAE ++ V+V+
 Sbjct: 705 EAKSPVVAKSPAEEK-SPAEEKPPAEAKSPAEEKSPAEEKSPAEEKSPAEEKSPA 762

Query: 168 PSTEETPDAAEATAVAE--NSVKVQPPPAEEA--PL-VEFPAEIQPPSAEE--SPSVELL 220
 P ++P E A ++AE + K + P EE P V+ P + + P EE SP
 Sbjct: 763 PEKAKSPVKEGAKSLAEAKSPEKAKSPVKEEIKPPAEVKSPEKAKSPMKEEAKSPEKAKT 822

Query: 221 AEILPPSAEESPSSEEP--PAEILPPPAEKSPSVELLGEIRSPSAQKAPIE-VQPLPAE-- 275
 ++ P A+ EE PA+I P KSP+ E E +SP ++ E V P E
 Sbjct: 823 LDVKSPEAKTPAKEEAKRPADIRSPQVKSPEAKE--EAKSPEKEETRTEKVAPKKEEVK 879

Query: 276 GALEEAPAKVEPPTVEETLAEVQPLPPEEAPREEARELQLSTAMETPAEEA-P-TEFQSP 333
 +EE AK P VEE E P P+ +E ++ A + AEE P TE
 Sbjct: 880 SPVEEVKAKEPPKKVEE--EKTPTPKTEVKESKKDEAPKEAQKPKAEKEEPLTEKPKD 936

Query: 334 LPKETTAEEASAEIQLLAATEPPADETPAE--ARSPLSEETSAAEEAHA-EVQSPLAEETT 390
 P E EEA + AA P +ETPA+ + + AE+A A E P +E
 Sbjct: 937 SPGEAKKEEAKK--KAAA--PEEETPAKLGVEEAKPKAEKAEKAEKAEKAEKAEK 991

Query: 391 A-EEASAEIQLLAIEAPADETPAEQSPLEETSAAEAPAEVQSPSA-KGVSIIEEAPLE 448
 EE A + E E+ + P + + EE Q PS K E++
 Sbjct: 992 KKEEVPAPEKKDKTEKTESKKPEEKPMQAKAEEDKGLPQEPSPKPKTEKAEKSSST 1051

Query: 449 LQPPSGEETTAEASAA 465
 Q S A E AA
 Sbjct: 1052 DQKDSQPSEKAPEDKAA 1068

Pedant information for DKFZphtes3_17f10, frame 3

Report for DKFZphtes3_17f10.3

[LENGTH]	710	
[MW]	75131.94	
[pI]	4.02	
[KW]	All_Alpha	
[KW]	LOW_COMPLEXITY	34.08 %

SEQ	MDRSQQTSTRTGYWTMMNIPPVEKVDKEQQTYFSESEIVVISRPDSSSTKSKEDALKHKSS
SEG
PRD	cc

SEQ	GKIFASEHPEFQPATNSNEEIGQKNISRTSFTQETKKGPPVLLEDELREEVTVPVQEGS
SEG
PRD	cc

SEQ	AVKKVASAEIEPPSTEFPAKIQPPLVEEATAKAEPRAEETHVQVQSTEETPDAAEAT
SEGxxxxxxxxxx
PRD	chhhhhhhcc

SEQ	AVAENSVKVQPPPAEAPLVEFPAEIQPPSAEESPSVELLAEILPPSAEESPSSEEPFAEI
SEG	xxxxx.....xx
PRD	hhhhhhcc

SEQ	LPPPAEKSPSVELLGEIRSPSAQKAPIEVQPLPAEGALEEAPAKVEPPTVEETLAEVQPL
SEG	xxxxxx.....xx
PRD	cc

SEQ	LPPEAPREEARELQLSTAMETPAEEAPTEFQSPLEKETTAEEASAEIQLLAATEPPADET
SEG	xx
PRD	ccccchhh

[illegible]

(No Prosite data available for DKFZphtes3 17f10.3)

(No Pfam data available for DKFZphtes3_17f10.3)

DKF2phtes3_17117

group: metabolism

DKF2phtes3_17117 encodes a novel 626 amino acid protein with similarity to transketolases (EC 2.2.1.1).

The novel protein contains a ATP/GTP-binding site motif A (P-loop). It is a new testis-specific transketolase. Transketolase requires thiamin pyrophosphate as cofactor and shows a wide specificity for both reactants, e.g. converts hydroxypyruvate and R-CHO into CO(2) and R-CHOH-CO-CH(2)OH.

The new protein can find application in modulation of metabolic pathways involving this transketolase activity and as a new enzyme for biotechnologic production processes.

strong similarity to transketolases

few EST hits (all from testis or pooled libraries containing testis)
testis specific transketolase?

Sequenced by GBF

Locus: unknown

Insert length: 2688 bp

Poly A stretch at pos. 2649, polyadenylation signal at pos. 2630

```

1  GACAAAAGAG AGATGATGGC CAACGACGCC AAGCCCGACG TGAAGACCGT
51  GCAGGTGCTG CGGGACACAG CCAACCGCCT GCGGATCCAT TCCATCAGGG
101 CCACGTGTGC CTCTGGTTCT GGCCAGCTCA CGTCGTGCTG CAGTGCAGCG
151 GAGGTCGTGT CTGTCTCTCT CTCCACACAG ATGAAGTATA AACAGACAGA
201 CCCAGAACAC CCGGACAAAC ACCGGTTCAT CCTCTCCAGG GGACATGCTG
251 CTCCTATCCT CTATGCTGCT TGGGTGGAGG TGGGTGACAT CAGTGAATCT
301 GACTTGCTGA ACCTGAGGAA ACTTCACAGC GACTTGGAGA GACACCCTAC
351 CCCGCCGATT CCGTTTGTG ACGTGGCAAC AGGGTCCCTA GGTGAGGGAT
401 TAGGTACTGC ATGTGGAATG GCTTATACTG GCAAGTACCT TGACAAGGCC
451 AGCTACCGGG TGTCTGCTT TATGGGAGAT GCGCAATCCT CAGAAGGCTC
501 TGTGTGGGAG GCTTTTGTCT TTGCCTCCCA CTACAACCTG GACAATCTCG
551 TGGCGGTCTT CGACGTGAAC CGCTTGGGAC AAAGTGGCCC TGCACCCTTT
601 GAGCATGGCG CAGACATCTA CCAGAATTGC TGTGAAGCCT TTGGATGGAA
651 TACTTACTTA GTGGATGGCC ATGATGTGGA GGCCTTGTGC CAAGCATTTT
701 GGCAAGCAAG TCAAGTGAAG AACAAAGCCTA CTGCTATAGT TGCCAAGACC
751 TTCAAAGGTC GGGGTATTCC AATATTGAG GATGCAGAAA ATTGGCATGG
801 AAAGCCAGTG CCAAAAGAAA GAGCAGATGC AATTGTCAAA TTAATTGAGA
851 GTCAGATACA GACCAATGAG AATCTCATAC CAAATCGCCC TGTGGAAGAC
901 TCACCTCAAA TAAGCATCAC AGATAATAAA ATGACCTCCC CACCTGCTTA
951 CAAAGTTGGT GACAAGATAG CTAATCAGAA AACATATGGT TTGGCTCTGG
1001 CTAAGTGGG CCGTGCAAAAT GAAAGAGTTA TTGTTCTGAG TGGTGACACG
1051 ATGAAGTCCA CCTTTTCTGA GATATTGAGG AAAGAACACC CTGAGCGTTT
1101 CATAGAGTGT ATTATTGCTG AACAAACAT GGAAGTGTG GCACTAGGCT
1151 GTGCTACACG TGGTCGAACC ATTGCTTTTG CTGGTGCTTT TGCTGCCTTT
1201 TTTACTAGAG CATTGATGTA GCTCCGAATG GGAGCCATTT CTCAAGCCAA
1251 TATCAACCTT ATTGGTTCCC ACTGTGGGGT ATCCACTGGA GAAGATGGAG
1301 TCTCCAGATG GGCCCTGGAG GATCTAGCCA TGTTCGGAAG CATTCCCAAT
1351 TGTACTGTTT TCTATCCAAG TGATGCCATC TCGACAGAGC ATGCTATTTA
1401 TCTAGCCGCC AATACCAAGG GAATGTGCTT CATTGCAACC AGCCAACCCG
1451 AAACGTCAGT TATTTATACC CCACAAGAAA ATTTTGAGAT TGGCCAGGCC
1501 AAGGTGGTCC GCCACGGTGT CAATGATAAA GTCACAGTAA TTGGAGCTGG
1551 AGTTACTCTC CATGAAGCCT TAGAAGCTGC TGACCATCTT TCTCAACAAG
1601 GTATTCTCTG CCGTGTCTAT GACCCATTTA CCATTAAACC CCTGGATGCC
1651 GCCACCATCA TCTCCAGTGC AAAAGCCACA GCGGCCGAG TTATCACAGT
1701 GGAGGATCAC TACAGGGAAG GTGGCATTGG AGAAGCTGTT TGTGCAGCTG
1751 TCTCCAGGGA GCCTGATATC CTGTTTCATC AACTGGCAGT GTCAGGAGTG
1801 CCTCAACGTG GGAAGACTAG TGAATTGCTG GATATGTTTG GAATCAGTAC
1851 CAGACACATT ATAGCAGCCG TAACACTTAC TTTAATGAAG TAACTAGGCC
1901 TTATTCTCAA AAAGTCAAGT CTATTGGCTT TGGCCCAAAA GCACCTGGAT
1951 CTTTGATATTA AATTCATGTT TATTGTCAAA AACCAATTAT TTATACCTAT
2001 ACAGTTGTAT TGTTCCTTTT AAAGCAAAGC CATTTAACAT CTTTCTTCAT
2051 TCCTAATTTG GAAATTAAAG TTTACCTTTC TGTTAATCTA TGTATAAATG
2101 TTACTCTGAG TTATTAATGT GGATTTTAAA ATTGTAAGCA ATAGAATAGG
2151 AAATAAAACA ACTACCTAAT ACAAAATATT CTGATAAGAC TACAAATATC
2201 TGACTGAGCT GGGGATTAAA GTAGAGGTAA CTGTATCTTA AATGAGTATG
2251 ATTTCTTTGT AAGTTAAAAA AATTGAAAT TAATTGTAGA CTTCAATAGT
2301 CCAAGTTTGG AAGGATGTTT GAGCTTTTGT ATAATGCCAT TTATACCTGC
2351 AGTTTTACAG ATAATGTTTG ACTGCAGTTG CCTTGGAAT TCCTCCAAAG
2401 TTTGCCCTCA TCTCTCTCT ACAGTTTGGG GGTGATGGTG CAGCAGTGGA
2451 ACATCTCTTG ATGCACCACA CTAATTGTGT TCTGTGAAGT GATGAAAGTA

```

2501 TAACTGGTTC TAGTTTGCAC ACTACACACA TAGTTTGTG AAGCTTCAGA
 2551 AATGTTTTTT CTTTCCTTG TGGCCAAACC AGTTTGTAA TCTGATTATA
 2601 TTCATCTGCT AATGATACTA AAGTTAATGT AATAAAGCAT TTAATAATCA
 2651 GAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA

BLAST Results

No BLAST result

Medline entries

96214928:
 Amplification of the transketolase gene in desensitization-resistant mutant
 Y1 mouse adrenocortical tumor cells.

99123875:
 Properties and functions of the thiamin diphosphate dependent enzyme transketolase.

Peptide information for frame 1

ORF from 13 bp to 1890 bp; peptide length: 626
 Category: strong similarity to known protein
 Classification: Metabolism
 Prosite motifs: ATP_GTP_A (595-603)

1 MMANDAKPDV KTVQVLRDTA NRLRIHSIRA TCASGSGQLT SCCSAAEVVS
 51 VLFFHTMKYK QTDPEHPDND RFILSRGHAA PILYAAWVEV GDISESDDLN
 101 LRKLHSDLER HPTPRLPFVD VATGSLGQGL GTACGMAYTG KYLDKASYRV
 151 FCLMGDGESE EGSVWEAFAP ASHYNLDNLV AVFDVNRLLGQ SGPAPLEHGA
 201 DIYQNCCEAF GWNTYLVDGH DVEALCQAFW QASQVKNKPT AIVAKTFKGR
 251 GIPNIEDAEN WHGKPVPKER ADAIVKLIES QIQTENLIP KSPVEDSPQI
 301 SITDIKMTSP PAYKVGDKIA TQKTYGLALA KLGRANERVI VLSGDTMNST
 351 FSEIFRKEHP ERFIECIIAE QNMVSVLGC ATRGRTIAFA GAFAAFFTRA
 401 FDQLRMGAIS QANINLIGSH CGVSTGEDGV SQMALEDLAM FRSIPNCTVF
 451 YPSDAISTEH AIYLAANTKG MCFIRTSQPE TAVIYTPQEN FEIGQAKVVR
 501 HGVNDKVTVI GAGVTLHEAL EAADHLSQQG ISVRVIDPFT IKPLDAATII
 551 SSAKATGGRV ITVEDHYREG GIGEAVCAAV SREPDILVHQ LAVSGVPQRG
 601 KTSSELLDMFG ISTRHIIAAV TLTLMK

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_17117, frame 1

SWISSPROT:TKT_MOUSE TRANSKETOLASE (EC 2.2.1.1) (TK) (P68).., N = 1,
 Score = 2222, P = 2.5e-230

SWISSPROT:TKT_RAT TRANSKETOLASE (EC 2.2.1.1) (TK).., N = 1, Score =
 2202, P = 3.3e-228

TREMBL:RN09256_1 product: "transketolase"; Rattus norvegicus
 Sprague-Dawley transketolase mRNA, complete cds., N = 1, Score = 2202,
 P = 3.3e-228

SWISSPROT:TKT_HUMAN TRANSKETOLASE (EC 2.2.1.1) (TK).., N = 1, Score =
 2200, P = 5.3e-228

>SWISSPROT:TKT_MOUSE TRANSKETOLASE (EC 2.2.1.1) (TK) (P68).
 Length = 623

HSPs:

Score = 2222 (333.4 bits), Expect = 2.5e-230, P = 2.5e-230
 Identities = 417/614 (67%), Positives = 501/614 (81%)

Query: 7 KPDVKTVQVLRDTANRLRIHSIRATCASGSGQLTSCCSAAEVVSVLFFHTMKYKQTDPEH 66

Sbjct: KPD + +Q L+DTANRLRI SI+AT A+GSG TSCCSAAE+++VLFFHTM+YK DP +
6 KPDQQKLQALKDTANRLRISSIQATTAAGSGHPTSCCSAAEIMAVLFFHTMRYKALDPRN 65

Query: 67 PDNDRFILSRGHAAPILYA AAWVEVGDISESDDLNLRLKLSDLERHPTPRLPFVDVATGSL 126
P NDRF+LS+GHAAPILYA W E G + E++LLNLRK+ SDL+ HP P+ F DVATGSL

Sbjct: 66 PHNDRFVLSKGHAAPILYAVWAEAGFLPEAELNLRLKISSDLGDHPVPKQAFSTDVATGSL 125

Query: 127 GQGLGTACGMAYTGKYLDKASYRVFCLMGDGESEGSVWEAFASHYNLDNLVAVFDVN 186
GQGLG ACGMAYTGKY DKASYRV+C++GDGE SEGSVWEA AFA Y LDNLVA+FD+N

Sbjct: 126 GQGLGACGMAYTGKYFDKASYRVYCMGLGDGEVSEGSVWEAMAFAGIYKLDNLVAIFDIN 185

Query: 187 RLQSGPAPLEHGADIYQNCCEAFGWNTYLVGDHDEALCQAFWQASQVKNKPTAIVAKT 246
RLQGS PAPL+H DIYQ CEAFGW+T +VDGH VE LC+AF QA K++PTAI+AKT

Sbjct: 186 RLQSDPAPLQHVDIYQKRCEAFGWHTIIVDGHVSVEELCKAFGQA---KHQPTAIIAKT 242

Query: 247 FKGRGIPNIEDAENWHGKPVPKERADAIVKLIESQIQTNENLIPKSPVEDSPQISITDIK 306
FKGRGI IED E WHGKP+PK A+ I++ I SQ+Q+ + ++ P ED+P + I +I+

Sbjct: 243 FKGRGITIEDKEAWHGKPLPKNMAEQIIQEIYSQVQSKKILATPPQEDAPSVDIANIR 302

Query: 307 MTSPPAYKVGDKIATQKTYGLALAKLGRANERVIVLSGDTMNSTFSEIFRKEHPERFIEC 366
M +PP+YKVGDKIAT+K YGLALAKLG A++R+I L GDT NSTFSE+F+KEHP+RFIEC

Sbjct: 303 MPTPPSYKVGDKIATRKA YGLALAKLGHASDRIALDGDTKNSTFSELFKEHPDRFIEC 362

Query: 367 IIAEQNMVSVAGCATRGRTIAFAGAFAAFFTRAFDQLRMGAISQANINLIGSHCGVSTG 426
IAEQNMVS+A+GCATR RT+ F FAAFFTRAFDQ+RM AIS++NINL GSHCGVS G

Sbjct: 363 YIAEQNMVSIAGCATRDRTVPFCSTFAAFFTRAFDQIRMAAISESNINLCGSHCGVSIG 422

Query: 427 EDGVSQMALEDLAFERSIPNCTVFYPSDAISTEHAIIYLAANTKGMCFIRTSQPETAVIYT 486
EDG SQMALEDLAFERS+P TVFYPSD ++TE A+ LAANTKG+CFIRTS+PE A+IY+

Sbjct: 423 EDGVSQMALEDLAFERSVPMSTVFYPSDGVATEKAVELAANTKGICFIRTSRPNIAIYS 482

Query: 487 PQENFEIGQAKVVRHGVNDKVTIVIGAGVTLHEALEAADHLSQQGISVRVIDPFTIKPLDA 546
E+F++GQAKVV +D+VTIVIGAGVTLHEAL AA+ L + IS+RV+DPFTIKPLD

Sbjct: 483 NNEDFQVGQAKVVLKSKDDQVTIVIGAGVTLHEALAAESLKKDKISIRVLDPFTIKPLDR 542

Query: 547 ATIISAKATGGRVITVEDHYREGGIGEA VCAAVSREPDLVHQLAVSGVPQRGKTSELL 606
I+ SA+AT GR++TVEDHY EGGIGEA V AAV EP + V +LAVS VP+ GK +ELL

Sbjct: 543 KLILDSARATKGRILTVEDHYEGGIGEA VSAAVVGEPGVTVTRLAVSQVPRSGKPAELL 602

Query: 607 DMFGISTRHIIAAV 620
MFGI I+ AV

Sbjct: 603 KMFGIDKDAIVQAV 616

Pedant information for DKFZphtes3_17117, frame 1

Report for DKFZphtes3_17117.1

[LENGTH] 626
[MW] 67877.52
[pI] 5.90
[HOMOL] SWISSPROT:TKT_MOUSE TRANSKETOLASE (EC 2.2.1.1) (TK) (P68). 0.0
[FUNCAT] m outer membrane and cell wall [M. jannaschii, MJ0681] 3e-48
[FUNCAT] g carbohydrate metabolism and transport [H. influenzae, HI1023] 9e-36
[FUNCAT] 01.05.01 carbohydrate utilization [S. cerevisiae, YPR074c] 5e-32
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YPR074c] 5e-32
[FUNCAT] 02.07 pentose-phosphate pathway [S. cerevisiae, YPR074c] 5e-32
[FUNCAT] 01.01.01 amino-acid biosynthesis [S. cerevisiae, YPR074c] 5e-32
[FUNCAT] i lipid metabolism [H. influenzae, HI1439] 3e-17
[FUNCAT] c energy conversion [H. influenzae, HI1233] 2e-09
[FUNCAT] 02.01 glycolysis [S. cerevisiae, YBR221c PDB1 - pyruvate dehydrogenase] 2e-05
[FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YBR221c PDB1 - pyruvate dehydrogenase] 2e-05
[BLOCKS] BL00801F
[BLOCKS] BL00801E
[BLOCKS] BL00801D Transketolase proteins
[BLOCKS] BL00801C Transketolase proteins
[BLOCKS] BL00801B Transketolase proteins
[BLOCKS] BL00801A Transketolase proteins
[SCOP] dltka2 3.28.1.2.1 Transketolase Transketolase, C-terminal domain 1e-21
[EC] 1.2.4.1 Pyruvate dehydrogenase (lipoamide) 8e-11
[EC] 1.2.4.4 3-Methyl-2-oxobutanoate dehydrogenase (lipoamide) 4e-10
[EC] 2.2.1.1 Transketolase 0.0
[EC] 2.2.1.3 Formaldehyde transketolase 1e-20
[PIRKW] transferase 0.0
[PIRKW] flavoprotein 2e-07
[PIRKW] Calvin cycle 1e-40
[PIRKW] heterotetramer 2e-07

```

[PIRKW]    pentose phosphate pathway 0.0
[PIRKW]    magnesium 1e-40
[PIRKW]    thiamine pyrophosphate 0.0
[PIRKW]    oxidoreductase 7e-12
[PIRKW]    fatty acid biosynthesis 4e-10
[PIRKW]    mitochondrion 2e-07
[PIRKW]    peroxisome 1e-20
[PIRKW]    homodimer 1e-40
[SUPFAM]   pyruvate dehydrogenase (lipoamide) alpha chain 1e-06
[SUPFAM]   pyruvate dehydrogenase (lipoamide) beta chain 7e-12
[SUPFAM]   ferredoxin 2[4Fe-4S]-related protein 8e-47
[SUPFAM]   thiamine pyrophosphate-binding domain homology 0.0
[SUPFAM]   pyruvate dehydrogenase (lipoamide) 6e-08
[SUPFAM]   ferredoxin 2[4Fe-4S] homology 8e-47
[SUPFAM]   hypothetical protein C2814 2e-21
[SUPFAM]   transketolase 0.0
[PROSITE]  ATP_GTP_A 1
[PFAM]     Transketolase
[KW]       Alpha_Beta
[KW]       3D
[KW]       LOW_COMPLEXITY 3.04 %

```

```

SEQ  MMANDAKPDVKTVQVLRDTANRLRIHSIRATCASGSGQLTSCCSAAEVVSVLFFHTMKYK
SEG  .....
lngsB .....HHHHHHHHHHHHCCCCHHHHHHHHHHHHHHHHHHHH-HHCCCT

SEQ  QTDPEHPDNDRLFILSRGHAAPILYAWEVVGDISEDLNLRKLHSDLERHPTPRLPFVD
SEG  .....
lngsB TTTTTTTTCEEEETTGGGHHHHHHHHHHHCTTCHHHHHHTTTTTTTTTTTTTTTTTTTC

SEQ  VATGSLGQGLGTACGMAYTGKYLKASYRVFCLMGDGESSEGSVWEAFASFASHYNLDNLV
SEG  .....
lngsB CCCCTTTTHHHHHHHHHHHHHHHHCBTTBTTEEECHHHHHCHHHHHHHHHHHHHHCTTTEE

SEQ  AVFDVNLRLGQSGPAPLEHGADIYQNCCEAFGWNTYLVLDGHDVEALCQAFWQASQVKNKPT
SEG  .....
lngsB EEEEECEETTEEAGGGCCCCCHHHHH-HHHCCEEEETTTTHHHHHHHHHHHHHHTTTTCE

SEQ  AIVAKTFKGRGIPNIEDAENWHGKVPKERADAIVKLIESQIQTNENLIPKSPVEDSPQI
SEG  .....
lngsB EEEECTTTTTTCHHHHHHHHHHHTCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCHHH

SEQ  SITDIKMTSPPAYKVGDKIATQKTYGLALAKLGRANERVIVLSGDTMNSTFSEIFRKEHP
SEG  .....
lngsB HHHHHHHHHTCCCTTTTTCBCHHHHHHHHHHHHHTTTTTTEEEETTTTHHHHCCTTCEECGG

SEQ  ERFIECIAEQNMVSVLGCATRGRTIAFAGAFAAFFTRAFDQLRMGAISQANINLIGSH
SEG  .....
lngsB GCEETTTTHHHHHHHHHHHHHTTTTEEEEGGGGGGGHHHHHHHHHHHCTTTEEEEC

SEQ  CGVSTGEDGVSQMALEDLAMFRSIPNCTVFYPSDAISTEHAIIYLAANTKGMCFIRTSQPE
SEG  .....
lngsB CCGGGTTTTTTTTTCHHHHHHHHCTTTTEEECCCHHHHHHHHHHHTTTTCEEEECCECCCB

SEQ  TAVIYTPQENFEIGQAKVVRHGVNDKVTVIGAGVTLHEALEAADHLSQQGISVRVIDPFT
SEG  .....
lngsB CCTTTTCHHHHHCC-CEEEETTTTTEEEECCHHHHHHHHHHHHHHHHHHCCCEEE...

SEQ  IKPLDAATIISAKATGGRVITVEDHYREGGIGEAACAASREPDILVHQLAVSGVPQRG
SEG  .....
lngsB .....

SEQ  KTSELDMFGISTRHIIAAVTLTLMK
SEG  .....
lngsB .....

```

Prosites for DKFZphtes3_17117.1

PS00017 595->603 ATP_GTP_A PDOC00017

Pfam for DKFZphtes3_17117.1

HMM_NAME Transketolase

HMM *vNtIRiLaMDAVEKANSNGHPGaPMGMAPMAHVWLqRMMRHNPNDPrWPN

Query	20	+N++RI ++ A + +SG +++++A++ VL+++M+++++DP P+ ANRLRIHSIRATCASGSGQLTSCCSAAEVVSVLFFHTMKYKQTDPEHPD	68
HMM		RDRFVLSNGHaCMLLYsMWHLYGYDMPMWDLkQFRQWHSrTPGHPEIGHT +DRF+LS GHA+++LY+ W + G +++++DL+++R++HS++ +HP ++	
Query	69	NDRFILSRGHAAPILYA+WVEVDG-ISESDLNLRKLHSDLERHPTPRLP	117
HMM		PGVEVTTGPLGQGIaNaVWMAIAERnLAATYNRPFGDI fDHYTYCFMGDG ++ +V+TG+LGQG++ +++++Y++++ D+++++C+MGDG	
Query	118	FV-DVATGSLGQGLG-----TACGMAYTGKYLDKASYRVFCLMGDG	157
HMM		CLMEGISWEACSLAGHMQLGNWiaFYDDNriSIDGdtIdWFqEDtYakRF + +EG++WEA ++A+H++L+N++A +D NR++++G++++ + D+Y+ +	
Query	158	ESSEGSVWEAFAFASHYNLDNLVAVFDVNRLGQSGPAPLEHGADIYQNC	207
HMM		EAYGWHVIEVENDGHDvEeIcaAIEeAKaekDRPTLIICRTVIGYGSPNk EA+GW++ +V DGHdVE++C A+ +A +K++PT+I ++T++G+G+PN	
Query	208	EAFGWNTYLV--DGHDVEALCQAFWQASQVKNKPTAIVAKTFKGRGIPNI	255
HMM		QGTHdWHGAPLGe* ++ + WHG+P +++	
Query	256	EDAENWHGKPVPE 269	
HMM		*PqWePnddkIATRKASQaLeaiGPaLPEfWGGsADLTPSNLTrWKGmv P++++ +DKIAT K+++ AL+++G A +++ +S+D+ +S+++++ ++	
Query	311	PAYKV-GDKIATQKTYGLALAKLGRANERVIVLSGDTMNSTFSEIFRKE	358
HMM		WFMPPSISTDCynGNWsGRYIHGIREHGMgAIMNGIALHGnFRPYGGT + + R+I++ I+E+M++++ G+A++G+ +++++ G	
Query	359	H-----PERFIECIIAEQNMVSVLGCATRGR-TIAFAGA	392
HMM		FMMFyDYARPAIRMAALMeLPVIWVWTHDSIGLGEDGPTHQPVEHLAHR F++F+++A++++RM A++ ++++++H++++ GEDG +++++E+LA+FR	
Query	393	FAAFFTRAFDQLRMGAISQANINLIGSHCGVSTGEDGVSQMALEDLAMFR	442
HMM		aIPNMsvVRPCDgNETayAWylAvERehTPtiLILSRQNLPLqErNPrqf +IPN +V++P+D+ T+ A YLA++++ +++++S ++ +++++ P +	
Query	443	SIPNCTVFYPSDAISTEHAIYLAANTKGM-CFIRTSQPETAVIYT-PQEN	490
HMM		ekvaRGYVLkDmdnePDVILIATGSEMELavaAAKILadEGikaRVVSM +++++++V + + + V++I++G++++A++AA+ L+ +GI +RV+++	
Query	491	FEIGQAKVVRHGVN--DKVTVIGAGVTLHEALEAADHLSQQGISVRVIDP	538
HMM		PCTeWFD.....kQDeEYReSVLPDhVPqRVaVEmGvtWCWYKYVGqg +++++D + +++++R +++DH++ ++++++V ++ +++ +	
Query	539	FTIKPLDAATIISAKATGGRVITVEDHYR-EGGIGEAVCAAVSREPDIL	587
HMM		GaIfGMNrFGESSGKAppevLYkMFGFTPENI* + +++ +++ ++ +L+ MFG+ +I	
Query	588	VHQLAVSGVPQR---GKTSELLDMFGISTRHI 616	

DKFZphtes3_17n12

group: transcription factors

DKFZphtes3_17n12.1 encodes a novel 804 amino acid protein which is nearly identical to mouse and trout SOX-LZ.

Sox proteins belong to the HMG box superfamily of DNA-binding proteins and are involved in the regulation of developmental processes as germ layer formation, organ development and cell type specification. Deletion or mutation of Sox proteins often results in developmental defects and congenital disease in humans. Sox proteins perform their function in a complex interplay with other transcription factors in a manner highly dependent on cell type and promoter context. The new protein is related to the SOX-LZ protein and contains an additional leucin-zipper.

The new protein can find application in modulating/blocking the expression of SOX-controlled genes.

nearly identical to mouse SOX-LZ

complete cDNA, complete cds, few EST hits
mouse and trout SOX-LZ, involved in spermatogenesis

Sequenced by GBF

Locus: unknown

Insert length: 2802 bp

Poly A stretch at pos. 2692, polyadenylation signal at pos. 2660

```

1 GGGATAGGAA AGATGAAAGG TCATGGTGAG CTTCAAGGAC ATGAAAGGTT
51 GTTGTCATCAT GTAACAATAG TAGATTGTTT TTTTCCTAA TATTTCTAGC
101 CAGCCCTCTAA GTCAGGTGAT GGAACAAATA CCTACAGTTT AGTCAGGTGA
151 AACAGGAGTG GGTGGAGGAA GGAAAGAAGA AAAATGGGAA GAATGTCTTC
201 CAAGCAAGCC ACCTCTCCAT TTGCTGTGTC AGCTGATGGA GAGGATGCAA
251 TGACCCAGGA TTAACTCTCA AGGGAAAAGG AAGAGGGCAG TGATCAACAT
301 GTGGCCTCCC ATCTGCCTCT GCACCCATA ATGCACAACA AACCTCACTG
351 TGAGGAGCTA CCAACACTTG TCAGTACCAT TCAACAAGAT GCTGACTGGG
401 ACAGCGTTCT GTCATCTCAG CAAAGAATGG AATCAGAGAA TAATAAGTTA
451 TGTTCCTTAT ATTCTTCCG AATACCTCT ACCTCACCAC ATAAGCCTGA
501 CGAAGGGAGT CGGGACCGTG AGATAATGAC CAGTGTACT TTTGGAACCC
551 CAGAGCGCCG CAAAGGGAGT CTTGCCGATG TGGTGGACAC ACTGAAACAG
601 AAGAAGCTTG AGGAAATGAC TCGGACTGAA CAAGAGGATT CCTCCTGCAT
651 GGAAAAACTA CTTTCAAAAG ATTGGAAGGA AAAAAATGGAA AGACTAAATA
701 CCAGTGAATC TCTTGAGAA ATTAAGGTA CACCTGAGAG CCTGGCAGAA
751 AAAGAACGGC AGCTCTCCAC CATGATTACC CAGCTGATCA GTTTACGGGA
801 GCAGCTACTG GCAGCGCATG ATGAACAGAA AAAACTGGCA CGCTCACAAA
851 TTGAGAAACA ACGGCAGCAA ATGGACCTTG CTCGCCAACA GCAAGAACAG
901 ATTGCCGAGC AACAGCAGCA ACTTCTGCAA CAGCAGCACA AAATTAATCT
951 CCTGCAGCAA CAGATCCAGG TTCAGGGTCA CATGCCTCCG CTCATGATCC
1001 CAATTTTTC ACATGACCAG CGGACTCTGG CAGCAGCTGC TGCTGCCCAA
1051 CAGGGATTCC TCTTCCCCC TGAATAACA TACAAACCA GTGATAACTA
1101 CCCCCTACAG TTCATTCCAT CAACAATGGC AGCTGCTGCT GCTTCTGGAC
1151 TCAGCCCTTT ACAGCTCCAG CAGCTCTATG CCGCTCAGCT GGCCAGCATG
1201 CAGGTGTCAC CTGGAGCAAA GATGCCATCA ACTCCACAGC CACCAACAC
1251 AGCAGGGACG GTCTCACCTA CTGGGATAAA AAATGAAAAG AGAGGGACCA
1301 GCCCTGTAA CTAAGTTAAG GATGAAGCAG CAGCACAGCC TCTGAATCTC
1351 TCATCCCGAC CCAAGACAGC AGAGCCTGTA AAGTCCCCAA CGTCTCCAC
1401 CCAGAACCTC TTCCAGCCA GCAAAACCAG CCCTGTCAAT CTGCCAAACA
1451 AAAGCAGCAT CCCTAGCCCC ATTGGAGGAA GCCTGGGAAG AGGATCCTCT
1501 TTAGGTAAAT GGAAAAGTCA ACACCAGGAA GAGACTTACG AATTAGATAT
1551 CCTATCTAGT CTCAACTCCC CTGCCCTTTT TGGGGATCAG GATACAGTGA
1601 TGAAGCCAT TCAGGAGGCG CGAAGATGC GAGAGCAGAT CCAGCGGGAG
1651 CAACAGCAGC AACAGCCACA TGGTGTGAC GGGAAACTGT CCTCCATAAA
1701 TAATATGGGG CTGAACAGCT GCAGGAATGA AAAGGAAAGA ACGCGCTTTG
1751 AGAATTTGGG GCCCCAGTTA ACGGGAAAGT CAAATGAAGA TGGAAAAC TG
1801 GGCCAGGTG TCATCGACCT TACTCGGCCA GAAGATGCAG AGGGAAGTAA
1851 AGCAATGAAT GGCTCTGCAG CTAACATACA GCAGTATTAT TGTGGCCAA
1901 CAGGAGGTGC CACTGTGGCT GAAGCACGAG TCTACAGGGA CGCCCGGGC
1951 CGTGCCAGCA GCGAGCCACA CATTAAGCGA CCAATGAATG CATTATGGT
2001 TTGGGCAAG GATGAGAGGA GAAAAATCCT TCAGGCCTTC CCCGACATGC
2051 ATAACCTCAA CATTAGCAA ATCTTAGGAT CTCGCTGGAA ATCAATGTCC
2101 AACCAGGAGA AGCAACCTTA TTATGAAGAG CAGGCCCGGC TAAGCAAGAT
2151 CCACCTTAGA AAGTACCCAA ACTATAAATA CAAACCCCGA CCGAAACGCA
2201 CCTGCATTGT CTGATGCAAA AAGCTTCGGA TTGGGGAGTA TAAGCAACTG
2251 ATGAGGTCTC GGAGACAGGA GATGAGGCAG TTCTTTACTG TGGGGCAACA
2301 GCCTCAGATT CCAATCACCA CAGGAACAGG TGTGTGTAT CCTGGTGCTA
2351 TCACTATGGC AACTACCACA CCATCGCCTC AGATGACATC TGACTGCTCT

```

```

2401 AGCACCTCGG CCAGCCCGGA GCCCAGCCTC CCGGTCATCC AGAGCACTTA
2451 TGGTATGAAG ACAGATGGCG GAAGCCTAGC TGGAAATGAA ATGATCAATG
2501 GAGAGGATGA AATGGAAATG TATGATGACT ATGAAGATGA CCCCAAATCA
2551 GACTATAGCA GTGAAATGA AGCCCCGAG GCTGTCAGTG CCAACTGAGG
2601 AGTTTTTGTG TGCTGAATTA AAGTACTCTG ACATTTTACC CCCCTCCCCA
2651 ACAAAGAGTT ATTAAAGAGC CCGCATGCAT TTGTGGCTCC ACAATTAAAA
2701 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2751 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2801 AA

```

BLAST Results

No BLAST result

Medline entries

95311974:

A gene that is related to SRY and is expressed in the testes encodes a leucine zipper-containing protein.

96032826:

The Sry-related HMG box-containing gene Sox6 is expressed in the adult testis and developing nervous system of the mouse.

Peptide information for frame 1

ORF from 184 bp to 2595 bp; peptide length: 804
Category: strong similarity to known protein

```

 1 MGRMSSKQAT SPFACAADGE DAMTQDLTSR EKEEGSDQHV ASHLPLHPIM
51 HNKPHSEELP TLVSTIQQDA DWDSVLSSQQ RMESENKLC SLYSFRNTST
101 SPHKPDEGSR DREIMTSVTF GTPERRKGS LADVVDTLKQ KLEEMTRTEQ
151 EDSSCMEKLL SKDWKEKMER LNTSELLGEI KGTPESLAEK ERQLSTMITQ
201 LISLREQLLA AHDEQKKLAA SQIEKQRQOM DLARQQEQEI ARQQOQLLQQ
251 QHKINLLQQQ IQVQGHMPP LMIPIPHDQR TLAIAAAQ GFLFPPGITY
301 KPGDNYPVQF IPSTMAAAAA SGLSPLQLQQ LYAAQLASMQ VSPGAKMPST
351 PQPPNTAGTV SPTGIKNEKR GTSPTVQVKD EAAQPLNLS SRPKTAEPVK
401 SPTSPTQNL FASKTSPVNL PNKSSIPSPI GGS LGRGSSL GKWKSQHQEE
451 TYELDILSSL NSPALFGDQD TVMKAIQEAR KMREIQREQ QQQPHGVVDG
501 KLSSINMGL NSCRNEKERT RFENLGPOLT GKSNEGDKLG PGVIDLTRPE
551 DAEGSKAMNG SAAKLQYYC WPTGGATVAE ARVYRDARGR ASSEPHIKRP
601 MNAFMVWAKD ERRKILQAFP DMHNSNISKI LGSRWKSMNS QEKQPYEEQ
651 ARLSKIHLEK YPNYKYKPRP KRTCIVDGKK LRIGEYKQLM RSRROEMRQF
701 FTVGQQPQIP ITTGTGVVYP GAITMATTPP SPQMTSDCSS TSASPEPSLP
751 VIQSTYGMKT DGGSLAGNEM INGEDEMEMY DDYEDDPKSD YSENEAPEA
801 VSAAN

```

BLASTP hits

Entry MMSOXL22_1 from database TREMBL:
product: "SOX-LZ"; Mouse mRNA for SOX-LZ, complete cds.
Score = 3910, P = 0.0e+00, identities = 764/801, positives = 774/801

Entry I51083 from database PIR:
SOX-LZ - rainbow trout
Score = 1774, P = 1.1e-287, identities = 365/532, positives = 431/532

Entry S59121 from database PIR:
SOX6 protein - mouse
Score = 2319, P = 1.2e-240, identities = 489/660, positives = 527/660

Entry AB006330_1 from database TREMBL:
gene: "mSox5L"; product: "SOX5"; Mus musculus mSox5L mRNA, complete cds.
Score = 1212, P = 8.9e-209, identities = 274/457, positives = 324/457

Entry MMU010604_1 from database TREMBL:
gene: "sox5"; product: "L-Sox5 protein"; Mus musculus mRNA for transcription factor L-Sox5
Score = 879, P = 4.2e-195, identities = 190/281, positives = 218/281

Alert BLASTP hits for DKFZphtes3_17n12, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_17n12, frame 1

Report for DKFZphtes3_17n12.1

[LENGTH] 804
 [MW] 89332.69
 [PI] 6.97
 [HOMOL] TREMBL:MMSOXLZ2_1 product: "SOX-LZ"; Mouse mRNA for SOX-LZ, complete cds. 0.0

[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YKL032c] 8e-07
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKL032c] 8e-07
 [FUNCAT] 01.07.07 regulation of vitamins, cofactors, and prosthetic groups [S. cerevisiae, YPR065w] 5e-06
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YBR089c-a] 7e-06
 [FUNCAT] 30.13 organization of chromosome structure [S. cerevisiae, YBR089c-a] 7e-06
 [FUNCAT] 03.01 cell growth [S. cerevisiae, YBR089c-a] 7e-06
 [FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YMR072w] 2e-04
 [FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YMR072w] 2e-04
 [SCOP] dlhmf_1.20.1.1.1 HMGl, fragments A and B [rat/hamster (Rattus 1e-13
 [SCOP] dllefa_1.20.1.1.6 Lymphoid enhancer-binding factor, LEF1 [mous 4e-15
 [SCOP] dlhrya_1.20.1.1.4 SRY [Human (Homo sapiens) 7e-17
 [PIRKW] DNA binding 4e-94
 [PIRKW] T-cell receptor 4e-07
 [PIRKW] leucine zipper 1e-38
 [PIRKW] alternative splicing 2e-07
 [PIRKW] transcription factor 4e-16
 [PIRKW] transcription regulation 1e-12
 [SUPFAM] HMG box homology 0.0
 [SUPFAM] unassigned HMG box proteins 4e-94
 [PROSITE] ATP_GTP_A 1
 [PROSITE] LEUCINE_ZIPPER 1
 [PROSITE] MYRISTYL 6
 [PROSITE] AMIDATION 1
 [PROSITE] CAMP_PHOSPHO_SITE 2
 [PROSITE] CK2_PHOSPHO_SITE 14
 [PROSITE] PKC_PHOSPHO_SITE 10
 [PROSITE] ASN_GLYCOSYLATION 6
 [PFAM] HMG (high mobility group) box
 [KW] Irregular
 [KW] 3D
 [KW] LOW_COMPLEXITY 13.81 %
 [KW] COILED_COIL 3.48 %

SEQ MGRMSSKQATSPFACAADGEDAMTQDLTSREKEEGSDQHVASHLPLHPIMHNKPHSEELP
 SEG
 COILS
 Inhm-

SEQ TLVSTIQDADWDVSLSSQQRMESENNKLCSLYSFRNTSTSPHKPDEGSRDREIMTSVTF
 SEG
 COILS
 Inhm-

SEQ GTPERRKGLADVVDTLKQKKLEEMTRTEQEDSSCKEKLKSKDWKEKMERLNTSELLGEI
 SEG
 COILS
 Inhm-

SEQ KGTPESLAEKERQLSTMITQLISLREQLLAHDEQKKLAASQIEKQRQOMDLARQQQEQT
 SEGXXXXXXXXXXXXXXXXXXXX
 COILSCCCCCCC
 Inhm-

SEQ ARQQQQQLLQQQHKKINLLQQQIQVQGHMPPLMIPFPHDQRTLAAAAAQQGFLFPFGITY
 SEGXXXXXXXXXXXXXXXXXXXX
 COILSCCCCCCCCCCCCCCCCCCCC
 Inhm-

SEQ KPGDNPVQVFIPTMAAAAAGLSPLQLQQLYAAQLASMQVSPGAKMPSTPQPNTAGTV
 SEGXXXXXXXXXXXXXXXXXXXX

```

COILS .....
lnhm- .....

SEQ  SPTGIKNEKRGTSPTVTQVKDEAAAQPLNLSSRPKTAEPVKSPTSPTQNLFPASKTSPVNL
SEG  .....
COILS .....
lnhm- .....

SEQ  PNKSSIPSPIGGSLGRGSSSLGKWKSQHQEETYELDILSSLNSPALFGDQDTVMKAIQEAR
SEG  ..xxxxxxxxxxxxxxxxxxxxx.....
COILS .....
lnhm- .....

SEQ  KMREQIQREQQQQPHGVGDKLSSINNMGNSCRNEKERTRFENLGPQLTGKSNEDGKLG
SEG  ..xxxxxxxxxxxxx.....
COILS .....
lnhm- .....

SEQ  PGVIDLTRPEDAEGSKAMNGSAAKLQQYYCWPTGGATVAEARVYRDARGRASSEPPIKRP
SEG  .....
COILS .....
lnhm- .....CCC

SEQ  MNAFMVWAKDERRKILQAFPMHNSNISKILGSRWKSMSNOEKQPYEEQARLSKIHLEK
SEG  .....X
COILS .....
lnhm- CCCHHHHHHHHHHHHHHHTTTTCCHHHHHHHHHHHTTTTTHHHHHHHHHHHHHHHHHHHHH

SEQ  YPNYKYKPRPKRTCIVDGKKLRIGEYKQLMRSRRQEMRQFFTVGQQPQIPITTTGTGVVYP
SEG  xxxxxxxxxxxx.....
COILS .....
lnhm- HHHTTTTTTT.....

SEQ  GAITMATTTTSPQMTSDCSSTSASPEPSLPVIQSTYGMKTDGGSLAGNEMINGEDEMEMY
SEG  .....xxxxxxx
COILS .....
lnhm- .....

SEQ  DDYEDDPKSDYSSENEAPAVSAN
SEG  xxxxxx.....
COILS .....
lnhm- .....

```

Prosites for DKFzphes3_17n12.1

PS00001	97->101	ASN_GLYCOSYLATION	PDOC00001
PS00001	172->176	ASN_GLYCOSYLATION	PDOC00001
PS00001	388->392	ASN_GLYCOSYLATION	PDOC00001
PS00001	422->426	ASN_GLYCOSYLATION	PDOC00001
PS00001	559->563	ASN_GLYCOSYLATION	PDOC00001
PS00001	626->630	ASN_GLYCOSYLATION	PDOC00001
PS00004	126->130	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	369->373	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	5->8	PKC_PHOSPHO_SITE	PDOC00005
PS00005	28->31	PKC_PHOSPHO_SITE	PDOC00005
PS00005	94->97	PKC_PHOSPHO_SITE	PDOC00005
PS00005	136->139	PKC_PHOSPHO_SITE	PDOC00005
PS00005	203->206	PKC_PHOSPHO_SITE	PDOC00005
PS00005	299->302	PKC_PHOSPHO_SITE	PDOC00005
PS00005	390->393	PKC_PHOSPHO_SITE	PDOC00005
PS00005	512->515	PKC_PHOSPHO_SITE	PDOC00005
PS00005	530->533	PKC_PHOSPHO_SITE	PDOC00005
PS00005	692->695	PKC_PHOSPHO_SITE	PDOC00005
PS00006	28->32	CK2_PHOSPHO_SITE	PDOC00006
PS00006	129->133	CK2_PHOSPHO_SITE	PDOC00006
PS00006	146->150	CK2_PHOSPHO_SITE	PDOC00006
PS00006	148->152	CK2_PHOSPHO_SITE	PDOC00006
PS00006	154->158	CK2_PHOSPHO_SITE	PDOC00006
PS00006	186->190	CK2_PHOSPHO_SITE	PDOC00006
PS00006	203->207	CK2_PHOSPHO_SITE	PDOC00006
PS00006	221->225	CK2_PHOSPHO_SITE	PDOC00006
PS00006	520->524	CK2_PHOSPHO_SITE	PDOC00006
PS00006	533->537	CK2_PHOSPHO_SITE	PDOC00006
PS00006	547->551	CK2_PHOSPHO_SITE	PDOC00006
PS00006	577->581	CK2_PHOSPHO_SITE	PDOC00006
PS00006	639->643	CK2_PHOSPHO_SITE	PDOC00006
PS00006	793->797	CK2_PHOSPHO_SITE	PDOC00006
PS00008	182->188	MYRISTYL	PDOC00008
PS00008	431->437	MYRISTYL	PDOC00008

PS00008	437->443	MYRISTYL	PDOC00008
PS00008	509->515	MYRISTYL	PDOC00008
PS00008	575->581	MYRISTYL	PDOC00008
PS00008	762->768	MYRISTYL	PDOC00008
PS00009	677->681	AMIDATION	PDOC00009
PS00017	526->534	ATP_GTP_A	PDOC00017
PS00029	187->209	LEUCINE_ZIPPER	PDOC00029

Pfam for DKFZphtes3_17n12.1

HMM_NAME	HMG (high mobility group) box		
HMM	*PKRPMNAYMLWMQEMReIKaENPNdMhNtEISKMiGEMWKnMsEEEEKm		
	+KRPMNA+M+W+++ R+KI + P DMHN++ISK++G +WK+MS +EK+		
Query	597	IKRPMNAFMVWAKDERRKILQAFF-DMHNSNISKILGSRWKSMSNQEQ	644
HMM	PYEdMAeeEKqRYMKEMPeYK*		
	PY+++ +++ + +++ +P+YK		
Query	645	PYYEEQARLSKIHLEKYPNYK	665

DKFZphtes3_17n18

group: intracellular transport and trafficking

DKFZphtes3_17n18 encodes a novel 782 amino acid protein with weak partial similarity to known proteins.

The novel protein contains a ATP/GTP-binding site motif A (P-loop) and a TonB-dependent receptor protein signature 1. In *E. coli*, the tonB protein interacts with outer membrane receptor proteins that mediate uptake of specific substrates into the periplasmic space. In the absence of tonB these receptors bind their substrates but do not carry out active transport. The novel protein seems to be involved in ATP-dependent transport of substances into the cell.

The new protein can find application in modulation of cell-permeability and transport of suitable substrates into the cell.

unknown receptor

protein contains TONB_DEPENDENT_REC_1 Pattern and ATP_GTP_A Pattern,

Sequenced by GBF

Locus: unknown

Insert length: 2853 bp

Poly A stretch at pos. 2806, no polyadenylation signal found

```

1  GTCCTTTTAA  GTCAGTAAAT  TGAACATAAGT  CGGTTATTCG  GCAAGCAGTT
51  CCTATAAAAA  ACTACATGGC  TAAGGTCTCT  AATGATTGAC  CACAAGCAGA
101  TCTTTTACCC  TCGGATCTCT  AGCTACAAAA  GGTCCCCACA  CTGAAGAAGC
151  CACTACCTCC  ACCACCACCA  GCACCACCAC  GTCCAGTGCT  GCTGGCAACC
201  ACTGGGGCAG  CCAAGCGCTC  CACCCCTCTCT  CCCACCATGG  CCCGTCAGGT
251  GCGCACCCAC  CAGGAGACCC  TGAACAGGTT  TCAGCAGCAG  TCCATCCACC
301  TGCTGACGGA  GCTCCTCAGA  CTGAAGATGA  AGGCCATGGT  GGAGTCTATG
351  TCGGTGGGTG  CCAACCCCTT  GGACATCACC  AGGCGCTTTG  TGGAGGCCAG
401  CCAGCTCCTC  CACCTCAATG  CCAAGGAGAT  GGCCTTCAAC  TGCCTGATCA
451  GCACAGCCCG  GAGAAGTGGC  TACAGCAGCG  GACAGTTGTG  GAAAGAGTCC
501  CTCGCAAAAC  TGTCCGCCAT  TGGGGTGAAC  TCGCCTTACC  AGCTGATCTA
551  CCACTCTTCC  ACAGCCTGTC  TGAGCTTTTC  TCTCTCTGCT  GGAAGAGAAAG
601  CCAAGAAGAA  AATAGGCAAA  TCTAGAACTA  CAGAAGATGT  CAGCATGCCG
651  CCCCTGCATC  GAGGAGTGGG  AACCCCTGCC  AACAGCCTGG  AGTTCAGCGA
701  CCCCTGCCCT  GAGGCCCGGG  AGAAGCTGCA  GGAGTTGTGT  CGCCACATAG
751  AAGCTGAAAG  GGCCACATGG  AAAGGGAGGA  ATATCTCCTA  CCCCATGATC
801  TTACGAAACT  ACAAGGCAAA  GATGCCCTCT  CATCTAATGT  TGGCCCGCAA
851  AGGAGACTCT  CAGACCCCGG  GTTTACATTA  CCCTCCCACT  CAGAGTGCTC
901  AGACTCTCAG  CCCCACCTCT  CACCCATCTT  CTGCCAACCA  TCATTTCAGT
951  CAGCATTGTC  AAGAGGGGAA  GGCACCCAAG  AAGGCCTTCA  AGTTTCATTA
1001  CACCTTCTAT  GATGGCTCCT  CCTTCGTTTA  CTATCCCTCT  GGAACGTCG
1051  CTGTATGTCA  GATCCCCACA  TGCTGCAGAG  GGAGAACCAT  CACCTGCCTC
1101  TTTAATGACA  TACCTGGATT  CTCCTTGCTG  GCCCTATTCA  ATACTGAAGG
1151  CCAGGGCTGT  GTTCACTACA  ACCTAAAAAC  CAGTTGCCCA  TATGCTTTAA
1201  TCTTGGATGA  GGAAGGTGGG  ACCACCAATG  ACCAGCAGGG  CTATGTAGTC
1251  CACAAGTGGC  GCTGGACTTC  CAGGACAGAG  ACCCTGCTTT  CCCTGGAATA
1301  CAAGGTGAAT  GAGGAAATGA  AACTAAAGGT  ACTGGGACAG  GACTCCATCA
1351  CAGTCACCTT  CACCTCCCTG  AATGAGACAG  TAACACTCAC  TGTGTCGGCC
1401  AACCAATTGT  CCCATGGAAT  GGCATATGAC  AAACGGCTGA  ACCGAGAAT
1451  CAGCAACATG  GACGACAAGG  TGTATAAGAT  GAGCCGAGCC  CTGGCTGAGA
1501  TCAAGAAGCG  GTTTCAGAAG  ACAGTGACTC  AGTTTCATTAA  TTCTATCTTG
1551  CTGGCCGCGC  GTCTGTTTAC  CATTGAATAT  CCCACCAAAA  AGGAGGAGGA
1601  AGAATTGTGT  CGGTTCAAGA  TGAGATCCAG  AACTCATCCC  GAGCGGCTCC
1651  CCAAGCTAAG  TTTATACTCA  GGAGAAAGTC  TTTTACGATC  TCAGTCAGGC
1701  CACCTGGAAT  CCTCAATTGC  AGAGACTTTG  AAGGATGAGC  CTGAGTCTGC
1751  TCCTGTGAGC  CCAGTTCGGA  AGACCACCAA  AATCCACACC  AAAGCCAAAG
1801  TCACATCCAG  AGGGAAGGCC  CGCGAGGGGC  GCAGCCCCAC  CAGGTGGGCG
1851  GCCTTGCCCT  CAGACTGCCC  GCTGGTGCTG  CGGAAGCTCA  TGCTCAAGGA
1901  AGACACCCGT  GCTGGCTGCA  AGTGCCTGGT  GAAGGCGCCC  CTGGTCTCTG
1951  ACGTGGAGCT  GGAGCGCTTC  CTGTTGGCGC  CCCGAGACCC  CAGCCAAAGT
2001  CTGGTGTTTG  GGATCATCTC  AAGCCAGAAC  TACACCAGCA  CTGGGCAGCT
2051  CCAGTGGCTG  CTGAACACTC  TCTACAACCA  CCAGCAGCGG  GGGCGTGGCT
2101  CCCCCTGCAT  CCAGTGCCGG  TATGACTCCT  ACCGCCTGCT  GCAGTATGAC
2151  CTGGACAGCC  CCCTGCAGGA  GGACCCTCCC  CTGATGGTGA  AGAAGAACTC
2201  TGTGGTGCAG  GGGATGATTC  TGATGTTTGC  CGGGGGGAAG  CTCATTTTTC
2251  GGGGCCGTGT  TTTGAATGGA  TATGGCCTCA  GCAAGCAGAA  TCTGCTGAAA
2301  CAGATCTTCC  GGTCTCAACA  GGATTACAAG  ATGGGCTACT  TCCTGCCGGA
2351  TGACTACAAA  TTCAGTGTTC  CCAACTCTGT  CCTGAGCCTG  GAGGATTCTG
2401  AATCAGTCAA  GAAAGCCGAG  TCAGAAGATA  TCCAAGGAAG  CAGCTCCTCA

```

```

2451 TTGGCCCTGG AAGACTATGT GGAÇAAGGAG TTATCTCTGG AGGCTGAGAA
2501 GACAAGAGAG CCTGAAGTGG AGCTACATCC TCTCAGCAGG GACAGCAAGA
2551 TAACTAGTTG GAAGAAGCAG GCCTCAAGA AGTAGCGCCA TCCTGGCAGC
2601 AGCCAAGTGA GCCAGGCCCC GCCCGGGGT GCTGGGGCTT CTTGCCAGCC
2651 CAGCCCTGCC TCCCCGGTCT CCCACCCTGT CCTCCAAGCT TCTATAATAA
2701 ACCAGCGGCG CTCCAGCATT GGGGTGAGGC TCTGGGGAAG GACAAAAAAA
2751 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAGGG
2801 CGGCCGAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAGGGCGG
2851 CCG

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 237 bp to 2582 bp; peptide length: 782

Category: putative protein

Prosite motifs: ATP_GTP_A (122-130)

TONB_DEPENDENT_REC_1 (1-44)

```

1 MARQVRTHQE TLNRFQQQSI HLLTELLRLK MKAMVESMSV GANPLDITRR
51 FVEASQLLHL NAKEMAFNCL ISTAGRSYGS SGQLWKESLA NMSAIGVNSP
101 YQLIYHSSSTA CLSFSLSAGK EAKKKGKSR TTEDVSMPPPL HRGVGTPANS
151 LEFSDPCPEA REKLQELCRH IEAERATWKG RNISYPMILR NYKAKMPSHL
201 MLARKGDSQT PGLHYPTAG AQTLSPTSHS SSANHHFSQH COEGKAPKKA
251 FKFHYTFYDG SSFVYYPGSG VAVCQIPTCC RGRITITCLFN DIPGFSLLAL
301 FNTGQGCYVH YNLKTSCPYV LILDEEGGT NDQQGYVVHK WSWTSRTETL
351 LSLEYKVNEE MKLKVLGQDS ITVTFTSLNE TVTLTVSANN CPHGMAYDKR
401 LNRRISNMDD KVKMSRALA EIKKRFQKTV TQFINSILLA AGLFTIEYPT
451 KKEEEEFVRF KMRSRTHPER LPKLSLYSGE SLLRSQSGHL ESSIATLKD
501 EPESAPVSPV RKTTKIHTKA KVTSRGKARE GRSPTRWAAL PSDCPLVLRK
551 LMLKEDTRAG CKCLVKAPLV SDVELERFLL APRDPSQVLV FGISSQNYT
601 STGQLQWLLN TLYNHQQRGR GSPCIQCRYD SYRLLQYDLD SPLQEDPPLM
651 VKKNSVVQGM ILMFAGGKLI FGGRVLNGYG LSKQNLKQI FRSQQDYKMG
701 YFLPDDYKFS VPNSVLSLED SESVKKAASE DIQSSSSSLA LEDYVEKELS
751 LEAEKTRERE VELHPLSRDS KITSWKQKAS KK

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_17n18, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_17n18, frame 3

Report for DKFZphtes3_17n18.3

```

[LENGTH]      782
[MW]           88030.16
[pI]           9.22
[BLOCKS]      BL00286 Squash family of serine protease inhibitors proteins
[PROSITE]     ATP_GTP_A      1
[PROSITE]     MYRISTYL      4
[PROSITE]     CAMP_PHOSPHO_SITE  3
[PROSITE]     CK2_PHOSPHO_SITE  14
[PROSITE]     PROKAR_LIPOPROTEIN  1
[PROSITE]     TONB_DEPENDENT_REC_1  1
[PROSITE]     PKC_PHOSPHO_SITE  10
[PROSITE]     ASN_GLYCOSYLATION  4
[KW]          Alpha_Beta

```

[illegible]

Prosite for DKFZphtes3 17n18.3

PS000001	91->95	ASN_GLYCOSYLATION	PDOC000001
PS000001	182->186	ASN_GLYCOSYLATION	PDOC000001
PS000001	379->383	ASN_GLYCOSYLATION	PDOC000001
PS000001	598->602	ASN_GLYCOSYLATION	PDOC000001
PS000004	403->407	CAMP_PHOSPHO_SITE	PDOC000004
PS000004	511->515	CAMP_PHOSPHO_SITE	PDOC000004
PS000004	652->656	CAMP_PHOSPHO_SITE	PDOC000004
PS000005	48->51	PKC_PHOSPHO_SITE	PDOC000005
PS000005	177->180	PKC_PHOSPHO_SITE	PDOC000005
PS000005	344->347	PKC_PHOSPHO_SITE	PDOC000005
PS000005	450->453	PKC_PHOSPHO_SITE	PDOC000005
PS000005	497->500	PKC_PHOSPHO_SITE	PDOC000005
PS000005	513->516	PKC_PHOSPHO_SITE	PDOC000005
PS000005	523->526	PKC_PHOSPHO_SITE	PDOC000005
PS000005	631->634	PKC_PHOSPHO_SITE	PDOC000005
PS000005	723->726	PKC_PHOSPHO_SITE	PDOC000005
PS000005	774->777	PKC_PHOSPHO_SITE	PDOC000005
PS000006	7->11	CK2_PHOSPHO_SITE	PDOC000006
PS000006	131->135	CK2_PHOSPHO_SITE	PDOC000006
PS000006	256->260	CK2_PHOSPHO_SITE	PDOC000006
PS000006	329->333	CK2_PHOSPHO_SITE	PDOC000006
PS000006	345->349	CK2_PHOSPHO_SITE	PDOC000006
PS000006	377->381	CK2_PHOSPHO_SITE	PDOC000006
PS000006	406->410	CK2_PHOSPHO_SITE	PDOC000006
PS000006	450->454	CK2_PHOSPHO_SITE	PDOC000006
PS000006	466->470	CK2_PHOSPHO_SITE	PDOC000006
PS000006	493->497	CK2_PHOSPHO_SITE	PDOC000006
PS000006	497->501	CK2_PHOSPHO_SITE	PDOC000006
PS000006	571->575	CK2_PHOSPHO_SITE	PDOC000006
PS000006	693->697	CK2_PHOSPHO_SITE	PDOC000006
PS000006	717->721	CK2_PHOSPHO_SITE	PDOC000006
PS000008	145->151	MYRISTYL	PDOC000008
PS000008	327->333	MYRISTYL	PDOC000008
PS000008	592->598	MYRISTYL	PDOC000008
PS000008	734->740	MYRISTYL	PDOC000008

PS00013	101->112	PROKAR_LIPOPROTEIN	PDOC00013
PS00017	122->130	ATP_GTP_A	PDOC00017
PS00430	1->44	TONB_DEPENDENT_REC_1	PDOC00354

(No Pfam data available for DKFZphtes3_17n18.3)

DKFZphtes3_18f3

group: testes derived

DKFZphtes3_18f3 encodes a novel 248 amino acid protein with partial similarity to human TNF-inducible protein CG12-1.

The novel protein contains two leucine zippers.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to TNF-inducible protein CG12-1

Sequenced by MediGenomix

Locus: unknown

Insert length: 4608 bp

Poly A stretch at pos. 4570, polyadenylation signal at pos. 4550

```

1  GACAGAAGTG AATGGGAATG GAGAGGCCGG CGGCCCGGGA GCCGCATGGG
51  CCCGACGCGC TCGCGCGCTT CCAGGGACTG CTGCTGGACC GCCGAGGCCG
101 GCTGCACCGC CAGGTGCTGC GCCTGCGCGA GGTGGCCCGG CGCCTGGAGC
151 GCCTGCGCAG GCGCTCCCTC GTAGCCAACG TGGCCGGCAG CTCGCTGAGC
201 GCAACGGGCG CCCTCGCCGC CATCGTGGGG CTCTCGCTCA GCCCGGTCAC
251 CCTGGGGACC TCGCTGCTGG TGTGGGCCGT GGGGCTGGGG GTGGCCACAG
301 CCGGAGGGGC CGTCACCATC ACGTCCGATC TCTCGCTGAT CTCTGCAAC
351 TCCCGGGAGC TCGGAGGGGT GCAGGAGATC GCGGCCACCT GCCAGGACCA
401 GATGCGAGAG ATCCTGAGCT GCCTCGAGTT TTTCTGCCGC TGGCAGGGCT
451 GCGGGGACCG CCAGCTGCTG CAGTGCGGGA GGAACGCCCT CATCGCCCTG
501 TACAATTCTG TCTACTTCAT CGTCTTCTTT GGCTCACGTG GCTTCCTCAT
551 CCCCAGGCGG GCGGAGGGGG ACACCAAGGT TAGCCAGGCC GTGCTGAAGG
601 CCAAGATTCA GAAACTGGCC GAGAGCCTGG AGTCTGAC CGGGGCTCTG
651 GACGAAGTCA GCGAGCAGCT GGAGTCTCGG GTTCAGCTCT GCACCAAGTC
701 CAGTCGTGGC CACGACCTCA AGATCTCTGC TGACCAGCGT GCAGGGCTGT
751 TTTTCTGAGA ACATCCTTTC CCCCTAATGA CCGAGGCCAG CAAATCATCC
801 TCATGGGATG CTCCAGAAAT TGTAGCTCCC TTAGGAAAAC ACCAAGCTGG
851 TTAGGAGGCC GAAGGCAAGG GATGAGAAAA ACTGTTTTTG AAGTGGGCAG
901 GTCCCCAAG CCCTTCTTTT CCCATCACTG TGACATCTGC CTGGGCTTGA
951 GTGCTACGGA CTTTTCAGTC TTCCTAGTGG AAAAATGTGA CCAAAAAACT
1001 CCTTTTCCCT TATCAAAAAA TTTCTGTCTA AACACAGCTG GGCAGGCACT
1051 CTGTTTTTAA AGTTATTTTC GGGTCCCTGA CCCTGCCCTG GTGGCTTGGC
1101 CTGAGACTGG AGAGAGTGCC ATCCTCTGGG TCCTCTCCAA GTCCTACTAG
1151 TCTTTGAAGT CCTCAAAATG TGCCTGAGGA AGGCATTGCG CTCTATTCCA
1201 GAATTTCTGA TACAAAGAAC TCCAGAATCC AGAGCAAATC AGCCCTTCTC
1251 TGAACGTTGT AGGATGGTTC AGAACCAGGA GAGGACCCTG GTGCTGATAT
1301 CTCCTCTCTT TCCCTTTCCC CTCAGCTTAC TTACTCCAG ATGCGGCCCTG
1351 GGTATGAAGT AGGCCCTTCC TGAGTGGCTC CCAATCCAGT CCTCCAAGTA
1401 CTCAGAGGGG AAGCCCGTGA AGCCGTCATC TAAGTCCTGC TCCCTCACAT
1451 GAAGCTGAGG GCCAGATAGA TGGAGCGACT GCCAATTCA TTTCCCGACA
1501 TCATTGTGTT CAGAAGAGAG TGATGGGTTT TGAGTTAGAC AGTCTGGGCG
1551 TTGAGACAGG CTTTGTCACT ACTGTGTGAG TGTAGCCACC TAATCTCTCT
1601 GAGACTGTGT AAAACAAAGA TGATAAAATC TCACCCTGTT GTGAGATATT
1651 AAATGAGCCA AAGTGCCTAG CATGATGGTG CTGGCTCATA TAGTGTAGTC
1701 CCTGGAATGG CAAATTAACA TCACCCAGGA ACTTGTTAGA AAGGCAAATT
1751 CTGGACACA ACCCTCCTGA TTTATGGAAT CAGAACTCTT GGCTGTGGGG
1801 CCCAGCAACC TGAGTTTAAA CAATTTCTCT GGGTGGTTCT GCGGCACACT
1851 AAGGTTTGAA AATCACTACA ACAAATGCTA ACTTCTAATC CCCTTGATGA
1901 GCTTTCACGA AGTCTCACGG CTCTCTAGG GACTCCATGG TCTTCAGAGT
1951 CGTTCACAGA TGACCAAGGA CAGACTGTGT CCCAGAAGCC AAAATGAGAG
2001 AGAGAGAGAG AGCACGCGTA CGTGACCCCT GGGGCAGTGT CTCACCGTAT
2051 GAATAAGGGA TGTAACACTA AAAGCCCATT AGGGGGCAGT GTTTCCCGCC
2101 TGTTGTAGAA ACTGGTACAG AAAGGATCCT ATATGAAGTT CCTGAACTG
2151 ACCTTTGTCT ATTATTACCT TCTCTGAAAA GTGCCAGTCC ATGTATTTTT
2201 TATTTATTTT AAGTTTGTA TTTAATTTT AATTATGTT TAGTGTGTC
2251 ATTTAATTTT ATTTAATCAC CACATTTAGA AAATAATAAG AGCAAGTTTC
2301 TAAATGGGAG ACTGCTGAGG CTCTTTGCAA GAGATGAGAT TAAGTTTGAG
2351 TTTCTAAGGC AGGGCATGAG CTGGAATAG CATTGCTTTC CTGTATGTG
2401 TCTCTCCTTC AGGGAGATT TTTTCTCTA GTGTTTAAAG TGATCCTTTG
2451 AAGTAAGTGT GGAGAGTCTT GAATGGCAAG ACCAGGAGCT GAGTTTAAGC
2501 TTGTAATGGA AGCTTGCAAT GTGGGATATA TAACTGAGGA AGCATATTTA
2551 TCCCTGAAGG ATTTTGCCAG AAGGTATCAC TTGACCTGGA AAAGGAATCT
2601 ATTTAGTTCA GGAAGATAA AAAGTTTAGA GGTATGTGAA GGAAGCACTT
2651 AGAAGTTGCA AGCCTGATGT CCTATCAAGT TATGCTTCTT GGGTGACAGA
2701 CAAATAGACT TGTCTTATGG TGGTGTATGG TTGCTATTTT ACTTTGGGGT

```



```

2751 CTGTAAGAAA CTGTCAGTGA AAATATGTAC AATTCCTTCA ATTTCCATTG
2801 TTAACAACTG TAATGTTGAA AAATAAGTTG AAAAGTCTTT GGGACCATAC
2851 ATGCAAAAAC GGTGCCTCTG TTACTTAATT ATTTAATATT CTATAAATGT
2901 ACCCAATCTG TCCGCACCCCT TCCAGTGAT GGGGCAGTAT GTCTGAGGAA
2951 GTATAATTTT AGTACTGGGG TCGGGGAGAG GAGGTGATGT TTCTACATTT
3001 TTATTTTTTC TATAAATTGC AATTGGTCTG TATGCTGGTT TATTTTGAAA
3051 TTTATATTGG TTTCTTTTCA AGCTGGTGTG ATCTCCTAGA CTGTTTCACC
3101 CAGATGCTAG CATTTTTTTT TTTTGTGAGA CAGAGTCTCA CTCTGTCACC
3151 TAGGCTGGAG TTGCAGTGGT TTGATCTCGG CTCAGTCAA CCTCCGACTC
3201 CTGGGTTCAC GCAATTCTTC TGCCTCAGCC TCCTGAGTAG CTGGGATTAC
3251 AGATGTGCAC CAGCACACCC GGCTAATTTT TTGTATTTT AGTAGAGACA
3301 GGGTTTCGCC ATGTTGGCCA GGCTGGTCTT GAACTCCTGG CCTTATGTGA
3351 TCCGCCACCC TTGGCTTCCC AAAGTGCTGG GATTACAGGC ATGAGCCACC
3401 TCGCCTGGCC AGATGCTAGC ATTTTAGATC AAACAATTCA TTTTAGATGA
3451 ATTGTTTTGT TTCACAATCA TTTTAAATCA TTTTAGAATG TACTTCACAT
3501 TATTAGTTGT GTTATGGCAT AAAGGTACAA CCATTCCTTA ACTCCATCTT
3551 TTATTAATGC TTAAGTTTAA ATTATATTCT TCCAATGCCT AAGCTATTCC
3601 CTAGAATTAA ACTGGGCACT TTTGGAAGCA GCAACAGTAA CAGCAGCAGC
3651 AAACTTTTCC TCTCATATT TGGGTGTATC AAAAGTTCTA GACTTTTGAA
3701 GTTATGATTT CAGTGGCCCA CTTTATTTCT AAGGAAGAGT GTCTACTTTG
3751 GAACGATACT TTGCACATAG TAGGAACCTA AGAAATACAT TTGAATAAAT
3801 ATAATTAAC TTTAGCTAT CTTAATGAGA ATTTGTTGAC AACAAAAGAT
3851 CATCCATCGC CTTATGTGTG AGTAAGATTG GAGCCTCTAT CAAGATTTAG
3901 TCAAGTTCAG TTAGATTGAT TCTAGAAACA AATATTTATT TCTTTCTTTT
3951 ACGGGGATGT GAATAAGGCT TTTCCCTAAG GCCTTCATT TTTAAACAAA
4001 CAGGTTGAAA TGGTATGTTG TAAAGAGAAA GACGGGAGAG AGGTATTTAG
4051 ATGATAAGTG TACTTCACAA AAATGCCAAA GTTGAAAAA TAGGTATGTT
4101 TGTTCTAAAT GTTTAAGTGC TTCTCTGTTA GGTTCGGGG CTGCAATCA
4151 TTTGAATTGT TCTGTTTCAC AATAAAGGAG ATTCAGTGGG TTCTGCATTT
4201 TCAGGATTCA ATAGAAGTGC TCCATTAAAA AAATAATCCT TAGCAAGCAT
4251 TCGAATCCTA ACTGCTTTGA TGCAGTTGCC CTCGGGCACC TGTCATTTC
4301 AATATGGTAG GTGTCAAAGT CAAAAGTATT TACTGGGAGA AAAAAAGAG
4351 GAGTGGTTGT AGAAGTCTCC CTAAATCAGA CATGTCAAGC AATCAGCCAA
4401 CGTGGTGTAT TTCTCATTCA ATATTTTAGT GTGAATTGAG AACTGAGAT
4451 AAAGACATCG TGCAGAGATA AATGGGGATA CAGTTAAATG TAGCAACTCT
4501 TGAGTTCATT TTTTCCCACT GTAGCAAAAT TAATGCTTTC TCTTTATTGA
4551 AATAAATTGC TCATTCCTCC AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
4601 AAAAAAGG

```

BLAST Results

Entry HSG27587 from database EMBL:
human STS SHGC-32548.
Score = 1951, P = 9.0e-101, identities = 411/425

Entry HS073350 from database EMBL:
human STS EST303564.
Score = 1417, P = 8.7e-58, identities = 285/287

Medline entries

No Medline entry

Peptide information for frame 2

ORF from the beginning to 580 bp; peptide length: 194
Category: questionable ORF
Classification: no clue

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_l8f3, frame 2

PIR:CGBOIS collagen alpha 1(I) chain - bovine (fragments), N = 1, Score
= 155, P = 4.5e-10

TREMBL:HSCG1PA1_1 gene: "COL1A1"; Human proalpha 1 (I) chain of type I
procollagen mRNA (partial)., N = 1, Score = 155, P = 6.5e-10

>PIR:CGBO1S collagen alpha 1(I) chain - bovine (fragments)
Length = 779

HSPs:

Score = 155 (23.3 bits), Expect = 4.5e-10, P = 4.5e-10
Identities = 60/152 (39%), Positives = 67/152 (44%)

Query: 7 GEAGGPGAAWARRAAALPGTAA--GPPRPAAPPGA--APARGGPAPGAPAQALPRSQRGR 62
G + G PG + AR PG GPP PA P GA AP G A A P SQ
Sbjct: 230 GDLGAPGPSGARGERGFPGERGVEGPPGAPGPRGANGAPGNDGAKGDAGAPGAPGSQGAP 289

Query: 63 QLAERNRPRRRHRGALAQPQHPGDLAAGVGRGAGGGHSRRGRHHVRSADLLQLPGAAE 122
L G P RGA PG GD +GA G + G VR L + PG A
Sbjct: 290 GL---QGMPEGE-RGAAGLPGPKGDRGDAGPKGADGAPGKDG----VRGLTGPIGPPGPAG 341

Query: 123 GAGDRGHL-P-GP-----DARDPELPRVFLPLAGLRGPPAA 156
GD+G P GP D +P P P AG GPP A
Sbjct: 342 APGDKGEAGPSGPAGTRGAPGDRGEPGPPG---P-AGFAGPPGA 381

Score = 121 (18.2 bits), Expect = 5.4e-05, P = 5.4e-05
Identities = 52/154 (33%), Positives = 60/154 (38%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARG-----GPAPGAPAQALPRSQRG 61
G G PGAA R P AGPP P P G ++G GPA G P + P G
Sbjct: 434 GATGFPGA--GRVGPGPSGNAGPPGPPGAPGKESKGRGETGPA-GRPGEVGP GPPG 491

Query: 62 RQLAERNRPRRRHRGALAQPQHPGDLAAGVGRGAGGGHSRRGRHHVRSADLLQLPGAA 121
A G P G PG PG RG G +RG R L PG +
Sbjct: 492 P--AGEKGAPGAD-GPAGAPGTGPGQGIAGQRGVVGLPGQRGE----RGFPGL---PGPS 541

Query: 122 EGAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAVRE 160
G +G R P P + GL GPP + RE
Sbjct: 542 GEPGKQGPSGASGERGPPGP---MGPPGLAGPPGESGRE 577

Score = 117 (17.6 bits), Expect = 1.8e-04, P = 1.8e-04
Identities = 52/148 (35%), Positives = 62/148 (41%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPA---PPGAAPARGGPAPGAPAQALPRSQRG-R 62
G G PG AR +A PG A G P A PPG + GP PG P A +G R
Sbjct: 416 GNVGAPGPKGARGSAGPPG-ATGFPGAAGRVGPPGPS-GNAGP-PGPPGAPGKESKGR 472

Query: 63 QLAERNRPRRRHRGALAQPQHPGDLAAGVGRGAGGGHSRRGRH--HHVRSADLLQLPGA 120
GRP G + PG PG GA G G + ++ LPG
Sbjct: 473 GETGPAGRP---GEVGP GPPGAPGAGADGPAGAPGTGPGQGIAGQRGVVGLPGQ 528

Query: 121 AEGAGDRGH--LPGPDARDPEL-PRVFLPLAGLRGPP 154
G+RG LPGP + P +G RGPP
Sbjct: 529 R---GERGFPGLPGPSGEPGKQGPS---GASGERGPP 559

Score = 117 (17.6 bits), Expect = 1.8e-04, P = 1.8e-04
Identities = 54/162 (33%), Positives = 64/162 (39%)

Query: 7 GEAGGPGAAWARRAAALPGT---AAGPPRPAAPPGAAPARG--GPA--PGAPAQALPRSQR 60
G G PG + PG A+GP P PPG G G A PG P + P +
Sbjct: 29 GPPGAPGPGQFGPPGEPGEPGASGPMGPRGPPGPKNGDDGAGKPRPGERGPPGPQ 88

Query: 61 G-RQLAERNRPR--RRHRGALAQPQHPGDLAAGVGRGAGGGHSRRGRHHV--RSLADLL 115
G R L G P + HRG G GD +G G G + R L
Sbjct: 89 GARGLPCTAGLPGMKHGRGFSGLDGAKGDAGPAGPKGEPGSPGENGAPQMGPRLPGFP 148

Query: 116 QLPGAA--EG-AGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAA 157
GAA G AG+RG +PGP P AG +GPP A
Sbjct: 149 GPKGAAGEPGKAGERG-VPGPPGAVG--PAGKDGEAGAQGPPGPA 190

Score = 113 (17.0 bits), Expect = 5.4e-04, P = 5.4e-04
Identities = 54/148 (36%), Positives = 58/148 (39%)

Query: 7 GEAGGPGAAWARRAAALPGTA-----AGPPRPAAP---PGAAPARGGPAP-GAPAQALPR 57
G AG PGA A PG A AGPP PA P PG G P P GA A P
Sbjct: 374 GFAGPPGADGQPGAKGEPGDAGAKGDAGPPGAPGAPGPPGPIGNVGA PGPKGARGSAGPP 433

Query: 58 SQRGRLAERNRPRRRHRGALAQPQHPGDLAAGVGRGAGGGHSRRGRHHVRSADLLQL 117
G A P G PG PG +G G GR V
Sbjct: 434 GATGFPGAAGRVGPPGPSGNAGPPGPPGAPGKESKGRGETGPAGRPGEVGP----- 486

Query: 118 PGAAEGAGDRGHLPGPD--ARDPELPRVFLPLAGLRG 152
PG AG++G PG D A P P +AG RG
Sbjct: 487 PGPPGPAGEKG-APGADGPAGAPGTGPP-QGIAGQRG 521

Score = 110 (16.5 bits), Expect = 1.3e-03, P = 1.2e-03

Identities = 54/151 (35%), Positives = 60/151 (39%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPP--AAPAR-GGPAP-GAPAQALPRSQR 62
 GE G G A + LPG A GPP A PG P G P P GA + +RG
 Sbjct: 194 GERGEQGPAGSPGFQGLPGPA-GPPGEAGKPGEQGVPGDLGAPGSGARGERGFPERGV 252

Query: 63 QLAERNRPRRRHARGALAQPGHPGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQLPGA 122
 + PR GA G GD A G+ G +G R A L PG
 Sbjct: 253 EGPPGPAGPRGANGAPGNDGAKGDAGAPGAPGSQGAAPGLQGMPE-RGAAGL---PGPK- 307

Query: 123 GAGDRGHLPGPDARD--PELPRVFLPLAGLRGPPAAA 157
 GDRG GP D P V L G GPP A
 Sbjct: 308 --GDRGDA-GPKGADGAPKDG V-RGLTGP IGP PGPA 340

Score = 109 (16.4 bits), Expect = 1.7e-03, P = 1.7e-03
 Identities = 55/154 (35%), Positives = 60/154 (38%)

Query: 4 NGN-GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARG-GPAPGAPAQALPRSQR 61
 NG+ GEAG PG R P A G P A PG RG GA A P +G
 Sbjct: 67 NGDDGEAGKPRP-GERGPPGPQAGRLPGTAGLPGMKHGRGFSGLDAGKGDAGPAGPKG 125

Query: 62 RQLAE-RNGRPRRRHARGALAQPGHPGDLAAGVGRGAGGGHSRRGRHHHVRSL----ADLL 115
 + NG P + G PG PG A G G G V A
 Sbjct: 126 EPGSPGENGAPGQ-MGPRGLPGFPKGAAGEPGKAGERGVPPGPAVGPAGKDGEAGAQ 184

Query: 116 QLPGAEEGAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAA 157
 PG A AG+RG GP A P F L G GPP A
 Sbjct: 185 GPPGPAGPAGERGE-QGP-AGSPG----FQGLPGPAGPPGEA 220

Score = 104 (15.6 bits), Expect = 6.6e-03, P = 6.6e-03
 Identities = 44/131 (33%), Positives = 49/131 (37%)

Query: 2 EVNGNGEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARGGPAP-GAPAQALPRSQR 60
 E GE G PG R LPG GP A PG A RG P P GA A +
 Sbjct: 126 EPGSPGENGAPGQMGP--GLPGFP-GPKGAAGEPGKAGERGVPPGPAVGPAGKDGEA 181

Query: 61 GRQLAERNRPRRRHARGALAQPGHPGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQLPGA 120
 G Q P RG G PG G+ G G G+ DL PG
 Sbjct: 182 GAQGP PG P AGPAGERGEQGPAGSPG--FQGLP-GPAGPPGEAGKPGEQGVPGDL-GAPGP 237

Query: 121 AEGAGDRGHLPG 132
 + G+RG PG
 Sbjct: 238 SGARGERG-FPG 248

Score = 104 (15.6 bits), Expect = 6.6e-03, P = 6.6e-03
 Identities = 43/131 (32%), Positives = 55/131 (41%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARGGPAPGAPAQALPRSQRQLAE 66
 GEAG G A R A PG G P P P G A GP PGA + + G A+
 Sbjct: 347 GEAGPSGPAGTRGA---PGDR-GEPPPGPAGFA---GP-PGADGQPGAKGEPGDAGAK 397

Query: 67 RNRGRPRRRHARGALAQPGHPGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQLPGAEEGAGD 126
 + P G PG G++ A +GA G G + A + PG + AG
 Sbjct: 398 GDAGPPGPAGPAGPPGPIGNVGAPGPKGARGSPGATGFPGA-AGRVGPPGPSGNAGP 456

Query: 127 RGHLPGPDARD 137
 G PGP ++
 Sbjct: 457 PGP-PGPAGKE 466

Score = 104 (15.6 bits), Expect = 6.6e-03, P = 6.6e-03
 Identities = 56/162 (34%), Positives = 62/162 (38%)

Query: 7 GEAGGPGAAWARRAAALPGTAA--GPPRPAAPPGAAPARGGPAPGAPAQALPRSQRQL 64
 G G PGA A G GP P P G A ARG P P Q PR +G
 Sbjct: 608 GPPGAPGAPGVGPAGKSGDRGETGPAGPIGPVGPAGARG---PAGP-QG-PRBKGZTG 662

Query: 65 AERNRPRRRHARG--ALAQP GHPGDLAAGVGRGAGGGHSRRGRHHHVRSLA-DLLQ-LPG 119
 + + + HRG PG PG GA G RG S D L LPG
 Sbjct: 663 ZZGBRGIKHGRGFSGLQGP GP GPSPGEQGPSGASGPAGPRGPPGSAGSPGKDGLNGLPG 722

Query: 120 AEGAGDRGHL--PGPDARDPELPRVFLPLAGLRGPPAAVREERLHRPVQ 168
 G RG GP A P P P G GPP+ L +P Q
 Sbjct: 723 PIGPPGPRGTGDAGP-AGPPGPPG---P-PGPPGPPSGGYDLSFLPQPPQ 768

Score = 101 (15.2 bits), Expect = 1.5e-02, P = 1.5e-02
 Identities = 49/148 (33%), Positives = 55/148 (37%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARGGPAPGAPAAA--QALPRSQRGR 62
 G AG PG A R PG A GP A G A A+G P P PA + P G
 Sbjct: 152 GAAGEPGKAGERGVPPGP-AGVP---AGKDGEAGAQPPGPAGPAGERGEQGPAGSPGF 207

Query: 63 QLAERNRPRRRHGALAQPGHFGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQLPGAEE 122
 Q P G + G PGDL A G G RG R + PG A
 Sbjct: 208 QGLPGPAGPPGEAGKPGEQGVPGDLGAP---GPSGARGERGFPGE-RGVEGP---PGPAG 260

Query: 123 GAGDRGHLPGPDARDPELPRVFLPLAGLRGPP 154
 G G PG D + P G +G P
 Sbjct: 261 PRGANG-APGNDGAKGDAGAPGAP--GSQGAP 289

Score = 100 (15.0 bits), Expect = 1.9e-02, P = 1.9e-02
 Identities = 40/130 (30%), Positives = 48/130 (36%)

Query: 7 GEAGGPGAAWARRAAALPGT--AAGPPRPAAPPGAAPARG--GPA--PGAPAQALPRSQR 60
 G G PG + PG A+G P PPG G G A PG P + P +
 Sbjct: 29 GPPGAPGPQGFQPPGEPGEPGASGPMGPRGPPGPGKNGDDGEAGKPRGPRGERGPPGPQ 88

Query: 61 G-RQLAERNRGP--RRHRGALAQPGHFGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQL 117
 G R L G P + HRG G GD +G G G + L
 Sbjct: 89 GARGLPGTAGLPGMKHGRGFSGLDGAKGDAGPAGPKGEPGSPGENGAPQMGPGRG-LPGF 147

Query: 118 PGAAEGAGDRG 128
 PG AG+ G
 Sbjct: 148 PGPKGAAGEPG 158

Score = 99 (14.9 bits), Expect = 2.5e-02, P = 2.5e-02
 Identities = 53/156 (33%), Positives = 61/156 (39%)

Query: 7 GEAGGPGAAWARRA---AALPGT--AAGPPRPAAPPGAAPARG--GPA---PGAPAQAL 55
 G G PGA R A PG A G P P P G + RG GPA P PA A
 Sbjct: 587 GRDGSPGAKGDRGETGPAGAPGPPGAPGAPGVPAGKSGDRGETGPAGPIGPVGPAGAR 646

Query: 56 -----PRSQGRQLAERNRPRRRHGALAQPGHFGDLAAGVGRGAGGGHSRRGRHHHV 108
 PR +G + + + HRG G PG + +G G G
 Sbjct: 647 GPAGPQGPGRBKGZTGZGZGBRGIKHGRGFSGLQGPPGPPSPGEGQPGSGASGPAGPRGP- 705

Query: 109 RSLADLLQLPGAEEGAGDRG--HLPDPDARDPELPRVFLPLAGLRGPP 154
 PG+A G G LPGP P PR AG GPP
 Sbjct: 706 -----PGSAGSPGKDGLNGLPGPIG--PPGPRGRTGDAGPAGPP 742

Score = 98 (14.7 bits), Expect = 3.3e-02, P = 3.3e-02
 Identities = 51/158 (32%), Positives = 58/158 (36%)

Query: 7 GEAGGPGAAWARRAAALPGTA-----AGPPRPAAPPGAAPARGGPAP-GAPAQALPRSQR 60
 G G G R AA LPG AGP PG RG P G P A +
 Sbjct: 287 GAPGLQMPGERGAAGLPGPKGDRGDAGPKGADGAPKDGVRGLTGPIGPPGAPAGPDK 346

Query: 61 GRQLAERNRPRRRHGA---LAQPGHFGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQL 117
 G A +G P RGA +PG PG GA G +G + D
 Sbjct: 347 GE--AGPSG-PAGTRGAPGDRGEPGPPGAPGAGPAGGADQPGAKGEPGDAGAKGDAGP- 402

Query: 118 PGAAEGAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAVR 159
 PG A AG G + A P+ R G G PAA R
 Sbjct: 403 PGFAGPAGPPGPIGNVGAAPGPKGARGSAGPPGATGFPPAAGR 444

Score = 96 (14.4 bits), Expect = 5.7e-02, P = 5.5e-02
 Identities = 46/152 (30%), Positives = 57/152 (37%)

Query: 6 NGEAGGPGAAWARRAAALPGTAA--GPPRPAAPPGAAPARGGPAPGAPA-QALPRSQRGR 62
 +G G PGA + PG G PA PG A G P P PA ++ R + G
 Sbjct: 574 SGREGAPGAEGSPGRDGSFGAKGDRGETGPAGAPGPPGAPGAPGVPAGKSGDRGETGP 633

Query: 63 QLAERNRPRRRHGALAQPGHFGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQLPGAEE 122
 P RG G G+ +G G RG H R + L PG
 Sbjct: 634 AGPIGPVGPAGARGPAGPQGPGRB-----KGZTGZGZGBRGIKH-RGFSGLQGPFGPPG 686

Query: 123 GAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAA 157
 G++G P A P AG RGPP +A
 Sbjct: 687 SPGEQG--PS-GASGP-----AGPRGPPGSA 709

Score = 94 (14.1 bits), Expect = 9.7e-02, P = 9.2e-02
 Identities = 45/134 (33%), Positives = 56/134 (41%)

Query: 24 PGTAAGPPRPAAPPGAAPARGGPA-PGAPAQALPRSQRQLAERNRPRRRH--GALAQ 80
 P G P P PG +G P PG P + P RG G P ++ G +
 Sbjct: 21 PSGPRGLPGPPGAPGPPQGFQPPGEPGEPGASGPMGPRGPP-----GPPGKNGDDGEAGK 75

Query: 81 PGHFGDLAA-GV--GRGAGGGHSRRGRHHHVRSLADLLQLPGAEEGAGDRGH--LPGPDA 135
 PG PG+ G RG G G H R + L G A AG +G PG +
 Sbjct: 76 PGRPGERGPPGPQAGRLPGTAGLPGMKGH-RGFSGLDGAKGDAGPAGPKGEPGSPGENG 134

Query: 136 RDEL-PRVFLPLAGLRGPPAAA 157
 ++ PR LP G GP AA

Sbjct: 135 APGQMGPGRG-LP--GFPGPKGAA 154

Score = 92 (13.8 bits), Expect = 1.7e-01, P = 1.5e-01
Identities = 52/155 (33%), Positives = 58/155 (37%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARGGP-APGAPAQALPRSQRGRQLA 65
GEAG G A R A G GPP PA G A G P A G P A + G
Sbjct: 347 GEAGPSGPAGTRGAPGDRGEP-GPPGPAGFAGPPGADGQPGAKGEPGDAGAKGDAGPPGP 405
Query: 66 ERNGRRRRHRGALAQPUGHGDLAAGVGRGAGGGHSRRGR--HHHVRSLADLLQLPGA-- 121
P G + PG G + GA G GR A PG A
Sbjct: 406 AGPAGPPGPIGNVGAPGPKGARGSAGPPGATGFPGAAGRVGPPGPSGNAGPPGPPGPAGK 465
Query: 122 EGA-GDRGHLPGPDARDPELPRVFLP-LAGLRGPPAA 156
EG+ G RG GP R E+ P AG +G P A
Sbjct: 466 EGSKGPRGET-GPAGRPGEVGPFGPPGAGEKGAPGA 501

Score = 92 (13.8 bits), Expect = 1.7e-01, P = 1.5e-01
Identities = 51/156 (32%), Positives = 57/156 (36%)

Query: 7 GEAGGPGAAWARRA---AALPGT--AAGPPRPAAPPGAAPARGGPAPGAPAQAL-PRSQR 60
G G PGA R A PG A G P P P G + RG P P + P R
Sbjct: 587 GRDGSFGAKGDRGETGPAGAPGPPGAPGVPVGPAGKSGDRGETGPAGPIGPVGPAGAR 646
Query: 61 GRQLAERNRPRRRHRGALAQPUGHGDLA-AGVG--RGAGGGHSRRGRH--HHVRSADLL 115
G A G PR +G + G G G +G G G A
Sbjct: 647 GP--AGPQG-PRGBKGZTGZGGRGKGRHGRFSGLGQPPGPPGSGPQSGASGPAGPR 703
Query: 116 QLPGAEGAGDRG--HLPGPDARDPELPRVFLPLAGLRGPP 154
PG+A G G LGGP P PR AG GPP
Sbjct: 704 GPPGSAGSPGKDLNGLPGPIG--PPGPRGRTGDAGPAGPP 742

Score = 90 (13.5 bits), Expect = 2.8e-01, P = 2.5e-01
Identities = 45/134 (33%), Positives = 53/134 (39%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARGGPAPGAPAQALPRSQRGRQ-LA 65
G G PG A + A G A P P P G A RG G P Q R +RG L
Sbjct: 485 GPPGPPGPAGEKGAPGADGPAGAPGTPG-PQGIAGQRG--VVGLPGQ--RGERGFPGLP 538
Query: 66 ERNGRRRRH--RGALAQPUGHGDLA----AGV---GR-GAGGGHSRRGRHHHVRSLADL 114
+G P + GA + G PG + AG GR GA G GR + D
Sbjct: 539 GPSGEPGKQGPSGASGERGPPGPMGPPGLAGPPGESGREGAPGAEGSPGRDGSFGAKGDR 598
Query: 115 LQL-PGAAEGAGDRGHLPGP 133
+ P A G PGP
Sbjct: 599 GETGPAGAPGPPGAPGAPGP 618

Score = 83 (12.5 bits), Expect = 1.8e+00, P = 8.3e-01
Identities = 49/156 (31%), Positives = 56/156 (35%)

Query: 7 GEAGGPGAAWARRAAALPGTAA--GPPRPAAPPGAAPARG--GPAP--GAPAQALPRSQR 60
G+AG GA A + G GPP PA PG G GPA GAP R +
Sbjct: 311 GDAGPKGADGAPGKDGVRGLTGPIGPPGAPAGPDKEAGPSGPAGTRGAPGD---RGE 367
Query: 61 GRQLAERNRPRRRHRGALAQPUGHGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQLPGA 120
G P G G PGD A G G G + ++ PG
Sbjct: 368 GPPGPAGFAGPPGADGQPGAKGEPGDAGAKGDAGPPGPAGPAGPPGPIGNVG----APGP 423
Query: 121 AEGAGDRGHLPGPDARDPELPRVFLP----LAGLRGPPAAVRE 160
G G PG RV P AG GPP A +E
Sbjct: 424 KGARGSAGP-PGATGFPGAAGRVGPPGPSGNAGPPGPPGPAGKE 466

Score = 82 (12.3 bits), Expect = 2.3e+00, P = 9.0e-01
Identities = 46/148 (31%), Positives = 52/148 (35%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARGGPAPGAPAQALPRSQRGRQLAE 66
G+AG PGA ++ A L G G A PG RG P A P R L
Sbjct: 275 GDAGAPGAPGSQGAPGLQGMPP-GERGAAGLPGPKGDRGDAGPKG-ADGAPGKDGVRGLTG 332
Query: 67 RNRPRRRHRGALAQPUGHGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQLPGAEGAGD 126
G P G PG G+ G G RG A PGA G
Sbjct: 333 PIGPP----GPAGAPGDKGEAGPSGPAGTRGAPGDRGEPGPPGP-AGFAGPPGADGQPGA 387
Query: 127 RGHLPGP-DARDPELPRVFLPLAGLRGPP 154
+G PG A+ P P AG GPP
Sbjct: 388 KGE-PGDAGAKGDAGPPG--P-AGPAGPP 412

Peptide information for frame 3

ORF from 12 bp to 755 bp; peptide length: 248
 Category: similarity to known protein
 Classification: unset
 Prosite motifs: LEUCINE_ZIPPER (17-39)
 LEUCINE_ZIPPER (24-46)

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_18f3, frame 3

TREMBL:AF070675_1 product: "TNF-inducible protein CG12-1"; Homo sapiens TNF-inducible protein CG12-1 mRNA, complete cds., N = 1, Score = 135, P = 1e-06

TREMBL:HS6802_1 gene: "dJ6802.1"; product: "dJ6802.1"; Homo sapiens DNA sequence from PAC 6802 on chromosome 22. Contains apolipoprotein L, myosin heavy chain, ESTs, CA repeat, STS and GSS., N = 1, Score = 107, P = 0.0023

>TREMBL:AF070675_1 product: "TNF-inducible protein CG12-1"; Homo sapiens TNF-inducible protein CG12-1 mRNA, complete cds.
 Length = 331

HSPs:

Score = 135 (20.3 bits), Expect = 1.0e-06, P = 1.0e-06
 Identities = 30/103 (29%), Positives = 55/103 (53%)

Query: 30 RLHRQVLRRLREVARRLRLRRSLVANVAGSSLSATGALAAIVGLSLSPVTLGTSLLVSA 89
 ++ + +LR +A +E + R ++NV SS A + ++ GL L+P T GTSL ++A
 Sbjct: 91 KIQESIEKLRALANGIEEVHRGCTISNVVSSSTGAASGIMSLAGLVLPFTAGTSLALTA 150

Query: 90 VGLGVATAGGAVTITS DL-SLIFCNSRELRRVQEI AATCQDQMR 132
 G+G+ A IT+ + + +S E + AT D+++
 Sbjct: 151 AGVGLGAASAVTGITTSIVEHSYTSSAEAE-ASRLTATSIDRLK 193

Pedant information for DKFZphtes3_18f3, frame 2

Report for DKFZphtes3_18f3.2

[LENGTH] 193
 [MW] 19708.24
 [pI] 11.90
 [KW] All_Alpha
 [KW] LOW_COMPLEXITY 55.44 %

SEQ TEVNGNGEAGGPGA AAWARRAAALPGTAAGPPRPAAPPGAAPARGGPAPGAPALPRSQR
 SEGXXX...
 PRD cccccccccccccchhhhhhhhhccccccccccccccccccccccccccccccccccccchhhhhh

SEQ GRQLAERNRPRRRRGALAQPGH PGDLAAGVGRGAGGGHSRRGRHHVRS LADLLQLPGA
 SEGXXX...
 PRD hhhhhhccchhhhhhhhhcccc

SEQ AEGAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAAVREERLHRPVQFCLLHRLWLW
 SEGXXX...
 PRD cchhhhhhhhhccccchhhhhhhhhc

SEQ LPHPQAGGGGHQ
 SEG xxxxxxxxxxxxxxx
 PRD ccccccccccccc

(No Prosite data available for DKFZphtes3_18f3.2)

(No Pfam data available for DKFZphtes3_18f3.2)

Pedant information for DKFZphtes3_18f3, frame 3

Report for DKFZphtes3_18f3.3

```

{LENGTH}      248
[MW]           27162.56
[pI]           9.92
[PROSITE]      LEUCINE_ZIPPER 2
[KW]           TRANSMEMBRANE 1
[KW]           LOW_COMPLEXITY 30.65 %
[KW]           COILED_COIL 12.10 %

SEQ  MGMPERPAAREPHGPDALRRFQGLLLDRRGRLHRQVLRRLREVARRLERLRRRSLVANVAGS
SEG  .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX.....
PRD  cccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcc
COILS .....
MEM  .....

SEQ  SLSATGALAAIVGLSLSPVTLGTSLLVSAVGLGVATAGGAVTITSDLSLIFCNSRELRRV
SEG  xxxxxxxxxxxx.....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.....
PRD  cchhhhhhhhhhhhhccccccccccccccccceeeceeeceeeceeeccchhhhhhh
COILS .....
MEM  .....MMMMMMMMMMMMMMMM.....

SEQ  QEIAATCQDQMREILSCLEFFCRWQCGDRQLLQCGRNASIALYNSVYFIVFFGSRGFLI
SEG  .....
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhccccchhhhhccccchhhhhceeeceeecccccc
COILS .....
MEM  .....

SEQ  PRRAEGDTKVSQAVLKAKIQKLAESLESCTGALDELSEQLSRVQLCTKSSRGHDLKISA
SEG  .....
PRD  cccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhcchhhhhhhhhhhhhhhhhhhccccceeehh
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCC.....
MEM  .....

SEQ  DQRAGLFF
SEG  .....
PRD  hhhhhccc
COILS .....
MEM  .....

```

Prosites for DKFZphtes3_18f3.3

```

PS00029      17->39  LEUCINE_ZIPPER      PDOC00029
PS00029      24->46  LEUCINE_ZIPPER      PDOC00029

```

(No Pfam data available for DKFZphtes3_18f3.3)

DKFZphtes3_1817

group: cell structure and motility

DKFZphtes3_1817 encodes a novel 1050 amino acid protein with weak partial similarity to ankyrins.

The novel protein contains an ATP/GTP-binding site motif A (P-loop) and an Ank repeat. Ankyrins are peripheral membrane proteins which interconnect integral proteins with the spectrin-based membrane skeleton. Thus the novel protein seems to be involved in coupling of cyto skeleton and cell membrane.

The new protein can find application in modulation of cyto skeleton-membrane interactions.

similarity to ankyrins

Sequenced by MediGenomix

Locus: unknown

Insert length: 4501 bp

Poly A stretch at pos. 4423, no polyadenylation signal found

```
1 GATCGCCGCG CGAGGGTGGT GGCATCGAG GTCCAGCAG CGGACGAGGG
51 AGGTGCCGCC GTCCGCCAGG ATGGGCTGGG AATGAAGCGA TGTAGCCTTT
101 TAAGAGATT GCTCTGACCC ATCTGAAGTC CATATGGCTC TGTATGATGA
151 AGACCTCTCTG AAAAATCCTT TCTATCTGGC TCTGCAAAAG TGCCGCCCTG
201 ACTTGTGCAG CAAAGTGCC CAAATCCATG GCATTGTCTT AGTACCCTGC
251 AAAGGAAGCC TGTCGAGCAG CATCCAGTCT ACTTGTCACT TTGAGTCCTA
301 CATTTTGATA CCTGTGGAAG AGCATTTCAT GACCTTAAAT GGAAGGATG
351 TCTTTATTCA AGGGAACAGG ATTAAATTAG GAGCTGGTTT TGCCTGTCTT
401 CTCTCAGTGC CCATTCTCTT TGAAGAACT TTCTACAATG AAAAAGAAGA
451 GAGTTTCAGC ATCCTGTGTA TAGCCCATCC TTTGGAAAAG AGAGAGAGTT
501 CAGAAAGACC TTTGGCACCC TCAGATCCCT TTTCCCTGAA AACCATTGAA
551 GATGTGAGAG AGTTCTTGGG AAGACACTCC GAGCGATTG ACAGGAACAT
601 CGCCTCTTTC CATCGAACAT TCCGAGAATG CGAGAGAAAG AGCCTCCGTC
651 ACCACATAGA CTCAGCGAAT GCTCTCTACA CCAAATGCCT CCAGCAGCTT
701 CTGAGGGAAT CTCACCTGAA AATGCTCGCC AAGCAGGAGG CCCAGATGAA
751 CCTGATGAAG CAGGCAGTGG AGATATACGT CCATCATGAA ATTTACAACC
801 TGATCTTTAA ATACGTGGGG ACCATGGAGG CAAGTGAGGA TGCGGCCTTT
851 AACAAATCA CAAGAAGCCT TCAAGATCTT CAGCAGAAAG ATATTGGTGT
901 GAAACCGGAG TTCAGCTTTA ACATACCTCG TGCCAAAAGA GAGCTGGCTC
951 AGCTGAACAA ATGCACCTCC CCACAGCAGA AGCTTGTCTG CTTGCGAAAA
1001 GTGGTGCGAG TCATTACACA GTCTCCAAGC CAGAGAGTGA ACCTGGAGAC
1051 CATGTGTGCT GATGATCTGC TATCAGTCCCT GTTATACTTG CTTGTGAAAA
1101 CGGAGATCCC TAATTGGATG GCAAATTTGA GTTACATCAA AAACCTCAGG
1151 TTTAGCAGCT TGGCAAAGGA TGAAGTGGGA TACTGCCTGA CCTCATTCGA
1201 AGCTGCCATT GAATATATTC GGCAAGGAAG CCTCTCTGCT AAACCCCTCG
1251 AGTCTGAGGG ATTTGGAGAC AGGCTGTTCC TTAAGCAGAG AATGAGCTTA
1301 CTCTCTCAGA TGACTTCGTC TCCCACCGAC TGCCCTGTTA AGCACATTGC
1351 ATCAGGTAAC CAGAAAGAAG TGGAGAGACT TCTGAGCCAA GAGGACCATG
1401 ATAAAGATAC CGTCCAAAAG ATGTGTCACC CTCTCTGCTT CTGCGATGAC
1451 TGTGAGAAAC TCGTCTCTGG GAGGTGAAT GATCCCTCAG TTGTCACTCC
1501 ATTCTCCAGA GACGACAGGG GGCACACCCC TCTCCATGTG GCTGCTGTCT
1551 GTGGGCAGGC ATCCCTCATC GACCTCCTGG TTTCCAAGGG CGCCATGGTA
1601 AATGCCACAG ACTACCATGG GGCACCTCCG CTCACCTGG CCTGTCAGAA
1651 GGGCTACCG AGCGTGACGC TGCTGCTGCT GCATACAAG CCCAGCGCGG
1701 AAGTGCAAGG CAACAATGGG AATACGCCAC TCCACCTGGC CTGCACCTAC
1751 GGCCACGAGG ACTGTGTGAA GGCTCTGGTT TACTACGACG TGGAGTCGTG
1801 CAGACTTGAC ATTGGCAATG AGAAAGGAGA CACCCCTCTA CACATTGCTG
1851 CCCGCTGGGG CTACCAAGGC GTCATAGAGA CATTGCTGCA GAACGGAGCG
1901 TCCACCGAGA TCCAGAACAG ACTGAAGGAG ACGCCCTCA AGTGTGCATT
1951 AAACCTCAAAG ATTCTGTCTG TAATGGAAGC CTATCACCTG TCCTTCGAGA
2001 GGAGGCAGAA GTCGTCCGAG GCCCTGTGTC AGTCCCGCCA CGCGTCCGTG
2051 GACTCCATCA GCCAAGAGTC CTCCACTTCC AGCTTCTCCT CCATGTGAGC
2101 CGGCTCAAGG CAGGAGGAGA CCAAGAAGGA CTACAGAGAG GTAGAAAAAC
2151 TTTTGAGAGC AGTTGCTGAT GGAGATCTAG AAATGGTGCG TTACCTGTTG
2201 GAATGGACAG AGGAGGACCT GGAGGATGCG GAGGACACTG TCAGTGAGCG
2251 AGACCCCGAA TTCTGTCAAC CGTTGTGCCA GTGCCCCAAG TGTGCCCCAG
2301 CTCAGAAGAG GCTGGCGAAG GTTCTTGCCA GTGGGCTTGG TGTGAACGTG
2351 ACCAGCCAGG ACGGCTCCTC CCCGCTGCAT GTCGCCGCCC TGCACGGCCG
2401 GCGGGACCTC ATCCGCCCTC TGCTGAAGCA CGGGGCCAAC GCAGGTGCCA
2451 GGAACGCAAG CCAAGCCGTC CCGCTCCACC TGGCCTGCCA CGAGGGCCAC
2501 TTTCAAGTGG TGAAGTGTCT GTTAGATTCC AATGCAAAAC CCAATAAGAA
2551 GGACCTCAGT GGAACACGCG CCTCATTTTA CGCCTGTCTC GGTGGCCATC
2601 ACGAGCTTGT GGCACGTGCT CTACAGCAGG GGGCCTCCAT TAACGCTTCT
2651 AACAAATAAG GCAACACAGC GCTGCAGGAG GCTGTGATTG AAAAGCAGCT
```



```

2701 CTTTCGTGGTA GAGCTGCTTC TGCTCCACGG AGCGTCAGTT CAGGTGCTGA
2751 ACAAGCGGCA GCGCACGGCT GTAGACTGTG CTGAACAGAA TTCAAAAATA
2801 ATGGAATTGC TTCAGGTGGT ACCAAGCTGT GTTGCTTCAT TAGATGATGT
2851 GGCTGAAACT GACCGCAAGG AGTATGTCAC TGTTAAGATC AGGAAAAAAT
2901 GGAACCTCAA ACTGTATGAT CTACCAGATG AGCCTTTTAC AAGACAGTTT
2951 TACTTTGTCC ACTCAGCTGG TCAGTTTAAG GGAAGACTT CRAAGGAGAT
3001 TATGGCAAGA GATAGAAGTG TCCCTAATTT AACCGAAGT TCTTTGCATG
3051 AGCCAGGGAG GCAAAGTGTC ACACTGAGAC AGAATAACCT GCCAGCTCAG
3101 AGTGGATCTC ATGCTGCTGA GAAAGGCAAC AGCGACTGGC CAGAGAGGCC
3151 TGGACTGACA CAGACTGGCC CTGGACACAG ACGGATGCTG CGGAGACACA
3201 CGGTAGAGGA TCGGTCGTG TCCCAGGGCC CGGAGGCTGC TGGCCCCCTC
3251 TCCACTCCCC AAGAGGTTAG TGCTTCCCGG TCCTAACAGG AATGAGGAGT
3301 TGTGTAACCC ACTGCTAGGA AGCAAGGATG CAACAAGATG ATGCTGAGCG
3351 TGAACACATC TGAGAACTAA ATGTGCTTCC ATGAGACTGG CTTGAGAAGT
3401 CTTCAGCACC AAGTTCCTGA AAGCTTTTCT GTGGCAGGAA AGAATGCAAC
3451 AAAAAAGTTA ACCACCACCA TCTCTCTCCT CTTCAAAGCT AATGAATACA
3501 ATTGAACAG ACAAATTC CAGTAGCATC CAGATCCTTA AGCCAGAGGT
3551 GCATGCTTCT TTTTAAGTAT GAGGGTTTGT TGGTCACAGT GGGAGAGGTT
3601 TCACCACCGC ATTCTGACCT CCTCCTCCCA AAAGGTGCTA AACCTCTCTG
3651 ACCTGTGTAC ATTCACAAAC CACAGCTAGA ATTCTCCAC CTAGGATTAA
3701 GCTGGAGAGA AGTAAGTAAT TTAGGTTTCA TGGTACTGTA GAGGCCAGGC
3751 TGAAATGTCA TATCTGAAGG AAGAAAGCAG CAGCTGGACA ATGTTTCTTT
3801 GCAAAGCAAC ACTCGAACCA AAAGATGCCT CAATCCCAT TGTATATTCA
3851 TTTTAGTGAA AGGATGCATC AGACCTGTT CACATCATGC ACATGGGAAA
3901 GGGTGGTTAT CATTTTCCTT CTAACAAGTA GGTACAGATA TTCGGTTACT
3951 ACACGTGCAC CTGTAGCAGT ATTCTAGAA ACATCCCTTT TTGTTGAGAA
4001 CCTCCCTTGA ATGCTGTGCA CACTCACACC TGACGGGATG GTTACTGGAT
4051 TAGAGAGTAG ATTTGGCACA TCTTTTCTTA GTCTTTTGT TCAAAATCAA
4101 AACTTAACAG CACAAACCAG GTCAGAGTTA CTTTCGGTTA GAATTTATTG
4151 CCATTTATTCT CTTTTTATAA ATTCTATAG ATTATACTGT TATTTTATG
4201 TTATTGGCCT AGAGCTACAC GTATATGGGT TTGTCCTGAG TCCGTTTCA
4251 AATGACCTTG TGATAGGGAA ATGGTTTGT CCATGTCTT GGAATACTT
4301 GTGTATGTAC AGAAGGAAGG GAGGGATTAT TTTTCTACAA AGTAATTTAT
4351 GATTTCTAAT TTTCTAATGT GCCTTGATA TGTGCCAAAT GATGGAAAAG
4401 AAACAGTAAA CTTTATGATT CTTAAAAAAA AAAAAAAGG AAAAAAAGG
4451 AAAAAAAGG AAAAAAAGG AAAAAAAGG AAAAAAAGG AAAAAAAGG
4501 G

```

BLAST Results

No BLAST result

*Medline entries

No Medline entry

Peptide information for frame 2

ORF from 134 bp to 3283 bp; peptide length: 1050
 Category: similarity to known protein
 Classification: Cell structure/motility
 Prosite motifs: ATP_GTP_A (945-953)

```

1  MALYDEDLK NPFYLAQKC RPDLCCKVAQ IHGIVLVPC KSLSSSIQST
51  CQFESYILIP VEEHFQTLNG KDVFIQGNRI KLGAGFACLL SVPILFEETF
101 YNEKEESFSI LCIAHPLEKR ESSEELAPS DPFSLKTIED VREFLGRHSE
151 RFDNRNIASFH RTFRECEKRS LRHHIDSANA LYTKCLQQLL RDSHLKMLAK
201 QEAQMNLMKQ AVEIYVHHEI YNLIFKYVGT MEASEDAAFN KITRSLQDLQ
251 QKDIGVKPEF SFNIPRAKRE LAQLNKCTSP QQKLVCLRV VQLITQSPSQ
301 RVNLETMCAD DLLSVLLYLL VKTEIPNWM NLSYIKNFRF SSLAKDELGY
351 CLTSFEAIE YIRQGSLSAK PPESEGFQDR LFLKQRMSSL SQMTSSPTDC
401 LFRHIASGNQ KEVERLLSQE DHDKDTVQKM CHPLCFCDCC EKLVSRLND
451 PSVVTFFSRD DRGHTPLHVA AVCGQASLID LLVSKGAMVN ATDYHGATPL
501 HLACQKGYQS VTLLLLHYKA SAEVQDNNGN TPLHLACTYG HEDCVKALVY
551 YDVESCRDI GNEKGDPLH IARWGYQGV IETLLQNGAS TEIQNRLKET
601 PLKCALNSKI LSVMEAYHLS FERRQKSEA PVQSPQRSVD SISQESSTSS
651 FSSMSAGSRQ EETKKDYREV EKLRLAVADG DLEMVRYLLE WTEEDLEDAE
701 DTVSAADPEF CHPLCQCPKC APAQKRLAKV PASGLGVNVT SQDGSPLHV
751 AALHGADLI RLLKHGANA GARNADQAVP LHLACQGHF QVVKCLDSN
801 AKPNKKDLGS NTPLIYACSG GHHELVALLL QHGASINASN NKGNTALHEA
851 VIEKHVFVVE LLLHGGASVQ VLNKRQRTAV DCAEQNSKIM ELLQVVPSCV

```

901 ASLDDVAETD RKEYVTVKIR KKWNSKLYDL PDEPFTTRQFY FVHSAGQFKG
 951 KTSREIMARD RSVPNLTEGS LHEPGRQSVT LRQNNLPAQS GSHAAEKGNS
 1001 DWPERPGLTQ TGPGRHRLR RHTVEDAVVS QGPEAAGPLS TPQEVASARS

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_1817, frame 2

TREMBL:HSU43965_1 gene: "ANK3"; product: "ankyrin G119"; Human ankyrin
 G119 (ANK3) mRNA, complete cds., N = 2, Score = 287, P = 3.7e-21

PIR:I49502 ankyrin - mouse, N = 3, Score = 365, P = 2.2e-27

TREMBL:HSANKY_2 product: "alt. ankyrin (variant 2.2)"; Human mRNA for
 ankyrin (variant 2.1), N = 2, Score = 380, P = 7.3e-31

SWISSPROT:ANK1_HUMAN ANKYRIN R (ANKYRINS 2.1 AND 2.2) (ERYTHROCYTE
 ANKYRIN)., N = 2, Score = 380, P = 8.2e-31

PIR:SJHUK ankyrin 1, erythrocyte splice form 1 - human, N = 2, Score =
 380, P = 8.2e-31

>TREMBL:HSANKY_2 product: "alt. ankyrin (variant 2.2)"; Human mRNA for
 ankyrin (variant 2.1)
 Length = 1,719

HSPs:

Score = 380 (57.0 bits), Expect = 7.3e-31, Sum P(2) = 7.3e-31
 Identities = 139/447 (31%), Positives = 207/447 (46%)

Query: 462 RGHTPLHVAACVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKAS 521
 +G+T LH+AA+ GQ ++ LV+ GA VNA G TPL+AA Q+ + V LL A+
 Sbjet: 77 KGNTALHIAALAGQDEVVRELVNYGANVNAQSQKGFPLYMAAQENHLEVVKFLENGAN 136

Query: 522 AEVDNNGNTPLHLACTYGHEDCVKALVYYDVES-CRL----- 558
 V +G TPL +A GHE+ V L+ Y + RL
 Sbjet: 137 QNVATEDGFTPLAVALQQGHENNVVAHLINYGTKGKVRPLALHIAARNDDTRTAAVLLQND 196

Query: 559 ---DIGNEKGDTPHLIAARWGYQGVETLLQNGASTEIQNRLKETPLKCALNSKILSVME 615
 D+ ++ G TPLHIAA + V + LL GAS + TPL A S+ +V+
 Sbjet: 197 PNPVLSKTGFTPLHIAAHYENLNVAQLLNLRGASVNFTPQNGITPLHIA--SRRGNVIM 254

Query: 616 AYHLSFERRQKSSEAPVQSPQRSVDSISQESSTS-SFSSMSAGSR-QEETKKDYREVEKL 673
 L +R + E + + ++ S + G+ Q +TK +
 Sbjet: 255 V-RLLLDRGAQI-ETKTKDELTPHCAARNGHVRISEILLDHGAPIQAKTKNGLSPIHM- 311

Query: 674 LRAVADGD-LEMVRYLLEWTEEDLEDAEDTVSAADPEFCHPLCQCPKCAPAQKRLAKVPA 732
 A GD L+ VR LL++ E ++D T+ P H C R+AKV
 Sbjet: 312 ---AAQGDHLDVCVRLLLQYDAE-IDDI--TLDHLP--LHVAHC----GHHRAKVL 358

Query: 733 S-GLGVNVTSDQGSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQQGHFQ 791
 G N + +G +PLH+A ++ LLLK GA+ A PLH+A GH
 Sbjet: 359 DKGAKPNSRALNGFTPLHIACKKNHVRVMEILLKTGASIDAVTESGLTPLHVASFMGHLP 418

Query: 792 VVKCLLDSNAKPNKKDLSGNTPLIYACSGGHELVALLQHGASINASNKGNLTALHEAV 851
 +VK LL A PN ++ TPL A GH E+ LLQ+ A +NA T LH A
 Sbjet: 419 IVKNLLQRGASPNVSNVKVETPLHMAARAGHTEVAKYLLQNKAKVNAKAKDDQTPLHCAA 478

Query: 852 IEKHVFVVELLLLHGASVQVLNKRQRTAVDCAEQNSKIMELLQVV 896
 H +V+LLL + A+ + T + A + + +L ++
 Sbjet: 479 RIGHTNMVKLLLENNANPNLATAGHTPLHIAAREGHVETVLALL 523

Score = 378 (56.7 bits), Expect = 1.2e-30, Sum P(2) = 1.2e-30
 Identities = 130/447 (29%), Positives = 195/447 (43%)

Query: 465 TPLHVAACVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKASAEV 524
 TPLH AA G + ++L+ GA + A +G +P+H+A Q + LLL Y A +
 Sbjet: 274 TPLHCAARNHVRRISEILLDHGAPIQAKTKNGLSPIHMAAQGDHLDVCVRLLLQYDAEIDD 333

Query: 525 QDNNGNTPLHLACTYGHEDCVKALVYYDVE-----SCR----- 557
 + TPLH+A GH K L+ + +C+
 Sbjet: 334 ITLDHLPPLHVAACHGHRVAKVLLDKGAKPNSRALNGFTPLHIACKKNHVRVMEILLKT 393

Query: 558 ---LDIGNEKGDTPHLIAARWGYQGVETLLQNGASTEIQNRLKETPLKCALNSKILSVME 614
 +D E G TPLH+A+ G+ +++ LLQ GAS + N ETPL A + V
 Sbjet: 394 GASIDAVTESGLTPLHVASFMGHLPVKNLLQRGASPNVSNVKVETPLHMAARAGHTEVA 453

Query: 615 EAYHLSFERRQKSSEAPVQSPQRSVDSISQESSTSSFSMSAGSRQEETKKDYREVEKLL 674
 + Y L + + + Q+P I + +A T L
 Sbjct: 454 K-YLLQNKAKVNAKAKDDQTPHCAARIGHTNMVKLLLENNANPNLATTAGH----TPLH 508

Query: 675 RAVADGDLEMVRYLLEWTEEDLEDAEDTVSAADPEFCHPLCQCPKCAPAQKRLAKVPASG 734
 A +G +E V LLE ++ A T P H + K A+ L +
 Sbjct: 509 IAAREGHVETVLALLE---KEASQACMTKKGFTP--LHVAARYGKVRVAELLER----D 559

Query: 735 LGVNVTSQDGSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQGHFQVVK 794
 N ++G +PLHVA H D+++LLL G + + + PLH+A +Q +V +
 Sbjct: 560 AHPNAGKNGLTPLHVAVHHNNLDIVKLLPRGGSPPHSPAWNNGYTPHIAAKQONQVEVAR 619

Query: 795 CLLDSNAKPNKKDLSGNTPLIYACSGGHELVALLQHGASINASNKGNLTALHEAVIEK 854
 LL N + + G TPL A GH E+VALL A+ N N G T LH E
 Sbjct: 620 SLLQYGGSSANAESVQGVTPHLAAQEGHAEMVALLSKQANGNLGNKSGLTPLHLVAQEG 679

Query: 855 HVFVVELLLLHGASVQVLNKRQRTAVDCAEQ--NSKIMELL 893
 HV V ++L+ HG V + T + A N K+++ L
 Sbjct: 680 HVPVADVLIKHGVMVDATTRMGYTPLHVASHYGNIKLVKFL 720

Score = 367 (55.1 bits), Expect = 1.8e-29, Sum P(2) = 1.8e-29
 Identities = 131/489 (26%), Positives = 210/489 (42%)

Query: 404 HIAS--GNQKEVERLLSQEDHDKDVTQKMCHPL-CFCDDCEKLVSGRLNDPSVVTFFSRD 460
 HIAS GN V LL + + + PL C + +S L D ++
 Sbjct: 244 HIASRRGNVIMVRLLDRAQIETKTDELTPLHCAARNGHVRISEILLDHGAPIQ-AKT 302

Query: 461 DRGHTPLHVAAVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKA 520
 G +P+H+AA + LL+ A ++ TPLH+A G+ V +LL A
 Sbjct: 303 KNGLSPIHMAAQGDHLDVRLLLQYDAEIDDITLDHLTPLHVAACHGHHRAKVLLDKGA 362

Query: 521 SAEVQDNNGNTPLHLACTYGHEDCVKALVYDVESCRLDIGNEKGDTPHIAARWGYQGV 580
 + NG TPLH+AC H ++ L+ +D E G TPLH+A+ G+ +
 Sbjct: 363 KPNSRALNGFTPLHIACKKNHVRVMELLK---TGASIDAVTESGLTPLHVASFHGLPI 419

Query: 581 IETLLQNGASTEIQNRLKETPLKCAL---NSKILSVMEAYHLSFERRQKSSEAPVQSPQR 637
 ++ LLQ GAS + N ETPL A ++++ + + K + P+ R
 Sbjct: 420 VKNLLQRGASPNVSNVVKVETPLHMAARAGHTEVAKYLLQNKAKVNAKAKDDQTPHCAAR 479

Query: 638 ----SVDSISQESSTSSFSMSAGSRQEETKKDYREVEKLLRAVADGDLEMVRYLLEWTE 693
 ++ + E++ + + +AG VE +L + + +T
 Sbjct: 480 IGHNTNMVKLLLENNANPNLATTAGHTPLHIAAREGHVETVLALLEKEASQACMTKKGFTP 539

Query: 694 EDLEDAEDTVSAAD---PEFCHPLCQ-----CP-KCAPAQKRLAKVPA---SGLGVNVTS 741
 + V A+ HP P A L V G + +
 Sbjct: 540 LHVAARYGKVRVAELLERDAHPNAGKNGLTPLHVAVHHNNLDIVKLLPRGGSPPHSPA 599

Query: 742 QDGSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQGHFQVVKCLLDSNA 801
 +G +PLH+AA + ++ R LL+G +A A + PLHLA Q+GH ++V LL A
 Sbjct: 600 WNGYTPLHIAAKQONQVEVARSLQYGGSSANAESVQGVTPHLAAQEGHAEMVALLSKQA 659

Query: 802 KPNKKDLSGNTPLIYACSGGHELVALLQHGASINASNKGNLTALHEAVIEKHVFVVEL 861
 N + SG TPL GH + +L++HG ++A+ G T LH A ++ +V+
 Sbjct: 660 NGNLGNKSGLTPLHLVAQEGHVPVADVLIKHGVMVDATTRMGYTPLHVASHYGNIKLVKF 719

Query: 862 LLLHGASVQVLNK 874
 LL H A V K
 Sbjct: 720 LLQHQAADVNAKTK 732

Score = 345 (51.8 bits), Expect = 4.2e-27, Sum P(2) = 4.2e-27
 Identities = 146/506 (28%), Positives = 233/506 (46%)

Query: 404 HIAS--GNQKEVERLLSQEDHDKDVTQK---MCHPLCFCDDCEKLVSGRLNDPSVVTFFS 458
 H+AS G+ K V LL +E + T +K H +++V +N + V +
 Sbjct: 50 HLASKEGHVKMVVELLHKEIILETTTKKGNTALHIAALAGQ-DEVVRELNVYGANVN--A 106

Query: 459 RDDRGHSTPLHVAAVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHY 518
 + +G TPL++AA ++ L+ GA N G TPL +A Q+G+++V L++Y
 Sbjct: 107 QSQKGFTPLYMAAQENHLEVVKFLENGANQNVATEDGFTPLAVALQOQHENVVAHLIN 166

Query: 519 KASAEVQDNNGNTPLHLACTYGHEDCVKALVYDVESCRLDIGNEKGDTPHIAARWGY 577
 +V+ P LH+A ++D A V + D+ ++ G TPLHIA +
 Sbjct: 167 GTKGKVR-----LPALHIAAR--NDDTRTAAVLLQNDP-NPDVLSKTGFTPLHIAAHYEN 218

Query: 578 QGVETLLQNGASTEIQNRLKETPLKCAL---NSKILSVMEAYHLSFERRQKSSEAPVQS 634
 V + LL GAS + TPL A N ++ ++ E + K P+
 Sbjct: 219 LNVAQLLNLRGASVNFTPNNGITPLHIASRRGNVIMVRLLDRAQIETKTDELTPLHC 278

Query: 635 PQRSVDSISQESSTSSFSMSAGSRQEETKKDYREVEKLLRAVADGD-LEMVRYLLEWTE 693
 R+ E + + A +TK + A GD L+ VR LL++
 Sbjct: 279 AARNGHVRISEILLDHGAPIQA-----KTKNGLSPIHM-----AAQGDHLDVRLLLQYDA 329

Query: 694 EDLEDAE-DTVSAAD-PEFC--HPLCQC-----PK-----CAPAQKRLAK 729
 E ++D D ++ C H + + P C R+ +
 Sbjct: 330 E-IDDITLDHLTPLHVAACHGHRVAKVLLDKGAKPNSRALNGFTPLHIACKKNHVRVME 388

Query: 730 VPA-SGLGVNVTSDQGSSPLHVAALHGRADLIRLLKKHGANAGARNADQAVPLHLACQGG 788
 + +G ++ ++ G +PLHVA+ G +++ LL+ GA+ N PLH+A + G
 Sbjct: 389 LLLKTGASIDAVTESGLTPLHVASFHGLPIVKNLLQRGASPNVSNVKVETPLHMAARAG 448

Query: 789 HFQVVKCLDSNAKPNKKDLSGNTPLIYACSGGHHELVALLLQH GASINASNNKGNTALH 848
 H +V K LL + AK N K TPL A GH +V LLL++ A+ N + G+T LH
 Sbjct: 449 HTEVAKYLLQNKAKVNAKAKDDQTPHCAARIGHTNMVKLLLENNANPNLATTAGHTPLH 508

Query: 849 EAVIEKHVFVVELLLH GASVQVLNKRQRTAVDCAEQNSKIM--ELL 893
 A E HV V LL AS + K+ T + A + K+ ELL
 Sbjct: 509 IAAREGHVETVLALLEKEASQACMTKKGFTPLHVAAKYGKVRVAELL 555

Score = 243 (36.5 bits), Expect = 1.6e-14, Sum P(2) = 1.6e-14
 Identities = 64/199 (32%), Positives = 97/199 (48%)

Query: 404 HIAS--GNQKEVERLLSQEDHDKDTVQKMCHPLCFCDCEKLVSGRLNDPSVVTFFSRDD 461
 H+A+ G + E LL ++ H + PL L +L P +P S
 Sbjct: 541 HVAAKYQKVRVAELLERDAHPNAAGKNGLTPLHVAVHHNNLDIVKLLPRGGSPPHSPAW 600

Query: 462 RGHTPLHVAACVQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLTLLHYKAS 521
 G+TPLH+AA Q + L+ G NA G TPLHLA Q+G+ + LLL +A+
 Sbjct: 601 NGYTPLHIAAKQNVQVEVARSLQYGGSSANAESVQGVTPHLHAAQEGHAEMVALLSSKQAN 660

Query: 522 AEVDNNGNTPLHLACTYGHEDCVKALVYDVESCRLDIGNEKGDTPHIAARWGYQGV 581
 + + +G TPLHL GH L+ + V +D G TPLH+A+ +G ++
 Sbjct: 661 GNLGNKSGLTPLHLVAQEGHVPVADVLIKHG---MVDATTRMGYTPLHVASHYGNIKLV 717

Query: 582 ETLLQNGASTEIQNRLKETPL 602
 + LLQ+ A + +L +PL
 Sbjct: 718 KFLLQHQAADVNAKTKLGYSPL 738

Score = 242 (36.3 bits), Expect = 5.0e-29, Sum P(2) = 5.0e-29
 Identities = 63/176 (35%), Positives = 92/176 (52%)

Query: 734 GLGVNVTSDQGSSPLHVAALHGRADLIRLLKKHGANAGARNADQAVPLHLACQGGHFQVV 793
 G VN T Q+G +PLH+A+ G ++RLLL GA + D+ PLH A + GH ++
 Sbjct: 229 GASVNFTPQNGITPLHIASRRGNVIMVRLLLDRGAQIETKTKDELTPHCAARNGHVRIS 288

Query: 794 KCLLDSNAKPNKKDLSGNTPLIYACSGGHHELVALLLQH GASINASNNKGNTALHEAVIE 853
 + LLD A K +G +P+ A G H + V LLLQ+ A I+ T LH A
 Sbjct: 289 EILLDHGAPIQAKTKNGLSPIHMAAQGDHLDVRLLLQYDAEIDDITLDHLTPLHVAACH 348

Query: 854 KHVFVVELLLHGA--SVQVLNKRQRTAVDCAEQNSKIMELLQVVPSCVASLDDVAET 909
 H V ++LL GA + + LN + C + + +MELL AS+D V E+
 Sbjct: 349 GHHRVAKVLLDKGAKPNSRALNGFTPLHIACKKNHVRVMEMLLLKTG---ASIDAVTES 403

Score = 242 (36.3 bits), Expect = 3.3e-14, Sum P(2) = 3.3e-14
 Identities = 80/284 (28%), Positives = 129/284 (45%)

Query: 404 HIAS--GNQKEVERLLSQEDHDKDTVQKMCHPLCFCDCEKLVSGRLNDPSVVTFFSRDD 461
 HIA+ G+ + V LL +E +K PL K+ L P +
 Sbjct: 508 HIAAREGHVETVLALLEKEASQACMTKKGFTPLHVAAKYQKVRVAELLERDAHPNAAGK 567

Query: 462 RGHTPLHVAACVQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLTLLHYKAS 521
 G TPLHVA ++ LL+ +G ++ +G TPLH+A ++ V LL Y S
 Sbjct: 568 NGLTPLHVAVHHNNLDIVKLLPRGGSPPHSPAWNGYTPLHIAAKQNVQVEVARSLQYGG 627

Query: 522 AEVDNNGNTPLHLACTYGHEDCVKALVYDVESCRLDIGNEKGDTPHIAARWGYQGV 581
 A + G TPLHLA GH + V L+ ++GN+ G TPLH+ A+ G+ V
 Sbjct: 628 ANAESVQGVTPHLHAAQEGHAEMVALLSSKQANG---NLGNKSGLTPLHLVAQEGHVPVA 684

Query: 582 ETLLQNGASTEIQNRLKETPLKCAL---NSKILSVMEAYHLSFERRQKSSEAPV-QSPQR 637
 + L+++G + R+ TPL A N K++ + + + K +P+ Q+ Q+
 Sbjct: 685 DVLIKHGVMVDATTRMGYTPLHVASHYGNIKLVKFLLQHQAADVNAKTKLGYSPLHQAQQ 744

Query: 638 S-VDSISQ--ESSTSSFSMSAGSRQEETK--DYREVEKLLRAVD 679
 D ++ ++ S S G+ K Y V +L+ V D
 Sbjct: 745 GHTDIVTLLKNGASPNEVSSDGTPLAIKRLGYISVTDVLKVVD 791

Score = 235 (35.3 bits), Expect = 7.9e-34, Sum P(2) = 7.9e-34
 Identities = 58/165 (35%), Positives = 83/165 (50%)

Query: 734 GLGVNVTSDQGSSPLHVAALHGRADLIRLLKKHGANAGARNADQAVPLHLACQGGHFQVV 793
 G N S G +PLH+AA G A+++ LLL AN N PLHL Q+GH V
 Sbjct: 625 GGSANAESVQGVTPHLHAAQEGHAEMVALLSSKQANGNLGNKSGLTPLHLVAQEGHVPVA 684

Query: 794 KCLLDSNAKPNKKDLSGNTPLIYACSGGHHELVALLQHGASINASNKNGNTALHEAVIE 853
 L+ + G TPL A G+ +LV LLQH A +NA G + LH+A +
 Sbjct: 685 DVLIKHGVMVDATTRMGYTPLHVASHYGNIKLVKFLQHQADVNKTKLGYSPHLHQAQQ 744

Query: 854 KHVFVVELLLLHGASVQVLNKRQRTAVDCAEQNS--KIMELLQVV 896
 H +V LLL +GAS ++ T + A++ + ++L+VV
 Sbjct: 745 GHTDIVTLLKNGASPNEVSSDGTTPLAIAKRLGYISVTDVLKVV 789

Score = 233 (35.0 bits), Expect = 7.9e-34, Sum P(2) = 7.9e-34
 Identities = 67/202 (33%), Positives = 100/202 (49%)

Query: 404 HIAS-GNQKEVERLLSQEDHDKDTVQKMCH--PLCFDDC-EKLVSGRLNDPSVVPFSR 459
 H+A+ G+ + RLL Q D + D + + H PL C V+ L D P SR
 Sbjct: 310 HMAAQGDHLDVCVRLLLQYDAEIDIT-LDHLTPLHVAACHGHRVAKVLLDKGA-KPNSR 367

Query: 460 DDRGHTPLHVAACVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYK 519
 G TPLH+A +++LL+ GA ++A G TPLH+A G+ + LL
 Sbjct: 368 ALNGFTPLHIAKKNHVRVMELLKTGASIDAVTESGLTPLHVASFMGHLPIVKNLLQRG 427

Query: 520 ASAEVQDNNNGNTPLHLACTYGHEDCVKALVYYDVESCRLDIGNEKGDTPHLHIAARWGYQG 579
 AS V + TPLH+A GH + K L+ +++ + TPLH AAR G+
 Sbjct: 428 ASPNSNVKAVETPLHMAARAGTEVAKYLLQ---NKAKVNAKAKDDQTPHCAARIGHTN 484

Query: 580 VIETLLQNGASTEIQNRKLPKCA 605
 +++ LL+N A+ + TPL A
 Sbjct: 485 MVKLLLENNANPNLATTAGHTPLHIA 510

Score = 226 (33.9 bits), Expect = 7.0e-33, Sum P(2) = 7.0e-33
 Identities = 53/153 (34%), Positives = 83/153 (54%)

Query: 743 DGSSPLHVAALHGRADLIRLLLLKHGANAGARNADQAVPLHLACQQGHFQVVKCLLDSNAK 802
 +G +PLH+AA + ++ R LL++G +A A + PLHLA Q+GH ++V LL A
 Sbjct: 601 NGYTPLHIAAKQNVARSLLQYGGSSANAESVQGVTPHLHAAQEGHAEMVALLSKQAN 660

Query: 803 PNKKDLSGNTPLIYACSGGHHELVALLQHGASINASNKNGNTALHEAVIEKHVFVVELL 862
 N + SG TPL GH + +L++HG ++A+ G T LH A ++ +V+ L
 Sbjct: 661 GNLGKNSGLTPLHLVAQEGHVPVADVLIKHGVMVDATTRMGYTPLHVASHYGNIKLVKFL 720

Query: 863 LLHGASVQVLNKRQRTAVDCAEQ--NSKIMELL 893
 L H A V K + + A Q ++ I+ LL
 Sbjct: 721 LQHQADVNAKTKLGYSPHLHQAQQGHTDIVTLL 753

Score = 198 (29.7 bits), Expect = 2.5e-11, Sum P(2) = 2.5e-11
 Identities = 51/157 (32%), Positives = 82/157 (52%)

Query: 737 VNVTSQDGSSPLHVAALHGRADLIRLLLLKHGANAGARNADQAVPLHLACQQGHFQVVKCL 796
 + T++ G++ LH+AAL G+ +++R L+ +GAN A++ PL++A Q+ H +VVK L
 Sbjct: 71 LETTTKNGNTALHIAALAGQDEVVRELNVYGANVNAQSQKGTPLYMAAQENHLEVVKFL 130

Query: 797 LDSNAKPNKKDLSGNTPLIYACSGGHHELVALLQHGASINASNKNGNTALHEAVIEKHV 856
 L++ A N G TPL A GH +VA L+ +G ALH A
 Sbjct: 131 LENGANQNVATEDGFTPLAVALQQGHENNVVAHLINYGTK---GKVRPLALHIAARNDDT 186

Query: 857 FVVELLLLHGASVQVLNKRQRTAVDCAE--QNSKIMELL 893
 +LL + + VL+K T + A +N + +LL
 Sbjct: 187 RTAAVLLQNDPNPDVLSKTGFTPLHIAAHYENLNVAQLL 225

Score = 186 (27.9 bits), Expect = 6.6e-29, Sum P(2) = 6.6e-29
 Identities = 55/143 (38%), Positives = 68/143 (47%)

Query: 463 GHTPLHVAACVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKASA 522
 GHTPLH+AA G + L+ K A G TPLH+A + G V LLL A
 Sbjct: 503 GHTPLHIAAREGHVETVLALLEKEASQACMTKKGFTPLHVAAYGKVRVAELLERDAHP 562

Query: 523 EVQDNNNGNTPLHLACTYGHEDCVKALVYYDVESCRLDIGNEKGDTPHLHIAARWGYQVIE 582
 NG TPLH+A + + D VK L+ S N G TPLHIAA+ V
 Sbjct: 563 NAAGKNGLTPLHVAVHHNLDIVKLLPRG-GSPHSPAWN--GYTPLHIAAKQNVQVEAR 619

Query: 583 TLLQNGASTEIQNRKLPKCA 605
 +LLQ G S ++ TPL A
 Sbjct: 620 SLLQYGGSSANAESVQGVTPHLA 642

Score = 182 (27.3 bits), Expect = 2.9e-28, Sum P(2) = 2.9e-28
 Identities = 54/185 (29%), Positives = 89/185 (48%)

Query: 738 NVTSQDGSSPLHVAALHGRADLIRLLLLKHGANAGARNADQAVPLHLACQQGHFQVVKCLL 797
 N+ ++ G +PLH+ A G + +L+KHG A PLH+A G+ ++VK LL
 Sbjct: 662 NLGNKSGLTPLHLVAQEGHVPVADVLIKHGVMVDATTRMGYTPLHVASHYGNIKLVKFL 721

Query: 798 DSNKPNKKDLSGNTPLIYACSGGHHELVALLQHGASINASNKNGNTALHEAVIEKHVF 857
 A N K G +PL A GH ++V LLL++GAS N ++ G T L A ++

Sbjct: 722 QHQADVNAKTKLGYSPHLHQAAQQGHTDIVTLLKNGASPNEVSSDGTTPLAIAKRLGYIS 781

Query: 858 VVELLLLHGASVQVLNKRQRTAVDCAEQNSKIMELLQVVPSCVASLDDVAETDRKEYVTV 917
V ++L + V ++ V + S P V + DV+E + +E ++

Sbjct: 782 VTDVLKV-----VTDETSFVLVSDKHRMS-----FPETVDEILDVSEDEGEELISF 827

Query: 918 KIRKK 922
K ++

Sbjct: 828 KAERR 832

Score = 180 (27.0 bits), Expect = 5.0e-29, Sum P(2) = 5.0e-29
Identities = 41/121 (33%), Positives = 67/121 (55%)

Query: 486 GAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKASAEVQDNNNGNTPLHLACTYGHEDCV 545
G +N + +G LHLA ++G+ + + LLH + E GNT LH+A G ++ V

Sbjct: 35 GVDINTCNQNLGLHLASKEGHVKMVELLHKEIILETTTKGNTALHIAALAGQDEVV 94

Query: 546 KALVYYDVESCRDLIGNEKGDTPHLIAARWGYQGVETLLQNGASTEIQNRLKETPLKCA 605
+ LV Y ++ ++KG TPL++AA+ + V++ LL+NGA+ + TPL A

Sbjct: 95 RELVNY---GANVNAQSQKGFPLYMAAQENHLEVVKFLENGANQNVATEDGFTPLAVA 151

Query: 606 L 606
L

Sbjct: 152 L 152

Score = 166 (24.9 bits), Expect = 3.4e-06, Sum P(2) = 3.4e-06
Identities = 89/318 (27%), Positives = 140/318 (44%)

Query: 448 LNDPSVVTFFSRDDRGTPLHVAAVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKG 507
L + + V ++DD+ TPLH AA G +++ LL+ A N G TPLH+A ++G

Sbjct: 457 LQNKAKVNAKAKDDQ--TPLHCAARIGHTNMVKLLLENNANPNLATTAGHTPLHIAAREG 514

Query: 508 YQSVTLLLLHYKASAEVQDNNNGNTPLHLACTYGHEDCVKALVYYD----- 552
+ L LL +AS G TPLH+A YG + L+ D

Sbjct: 515 HVETVLALLEKEASQACMTKKGFTPLHVAAKYKVRVAELLERDAHFNAAAGKNGLTPLH 574

Query: 553 --VESCRLDI-----GNE-----KGDTPHLIAARWGYQGVETLLQNGASTEIQNRL 597
V LDI G+ G TPLHIAA+ V +LLQ G S ++

Sbjct: 575 VAVHHNNLDIVKLLPRGGSPHSPAWNGYTPLHIAAKQNVQVEVARSLQYGGSSANAESVQ 634

Query: 598 KETPLKCALNSKILSVMEAYHLSFERRQKSSEAPVQSPQRSVDSISQESSTSSFSM-SA 656
TPL A M A LS +Q + +S + ++QE +

Sbjct: 635 GVTPLHLAAQEGHAE-MVALLS---KQANGNLGNKSGLTPLHLVAQEGHVPVADVLIKH 690

Query: 657 GSRQEETKKDYREVEKLLRAVADGDLEMYRYLLEWTEEDLEDAEDTVSAADPEFCHPLCQ 716
G + T + L A G+++V++LL+ + D+ +A+ + + PL Q

Sbjct: 691 GVMVDATTR--MGYTPLHVASHYGNIKLVKFLQH-QADV-NAKTKLGY-----PLHQ 740

Query: 717 CPKCAPAQKRLAKVPASGLGVNVTSDGSSPLHVA 751
+ + + +G N S DG++PL +A

Sbjct: 741 AAQQGHTDI-VTLLKNGASPNEVSSDGTTPLAIA 774

Score = 162 (24.3 bits), Expect = 1.8e-07, Sum P(2) = 1.8e-07
Identities = 48/149 (32%), Positives = 71/149 (47%)

Query: 737 VNVTSQDGSSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQGGHFQVVKCL 796
V D ++ AA G D L++G + N + LHLA ++GH ++V L

Sbjct: 5 VGFREDAATSFLRAARSGNLDKALDHLRNGVDINTCNQNLGLHLASKEGHVKMVEL 64

Query: 797 LDSNAKPNKKDLGNTPLIYACSGGHELVALLQH GASINASNNGNTALHEAVIEKHV 856
L GNT L A G E+V L+ +GA++NA + KG T L+ A E H+

Sbjct: 65 LKHEIILETTTKGNTALHIAALAGQDEVVRELVNYGANVNAQSQKGFPLYMAAQENHL 124

Query: 857 FVVELLLLHGASVQVLNKRQRTAVDCAEQ 885
VV+ LL +GA+ V + T + A Q

Sbjct: 125 EVVKFLENGANQNVATEDGFTPLAVALQ 153

Score = 158 (23.7 bits), Expect = 5.7e-26, Sum P(2) = 5.7e-26
Identities = 38/135 (28%), Positives = 65/135 (48%)

Query: 460 DDRGHTPLHVAAVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYK 519
+ G LH+A+ G ++ L+ K ++ T G T LH+A G V L++Y

Sbjct: 42 NQNLGLHLASKEGHVKMVELLHKEIILETTTKGNTALHIAALAGQDEVVRELVNYG 101

Query: 520 ASAEVQDNNNGNTPLHLACTYGHEDCVKALVYYDVESCRDLIGNEKGDTPHLIAARWGYQ 579
A+ Q G TPL++A H + VK L+ ++ E G TPL +A + G++

Sbjct: 102 ANVNAQSQKGFPLYMAAQENHLEVVKFLE---NGANQNVATEDGFTPLAVALQGHEN 158

Query: 580 VIETLLQNGASTEIQ 594
V+ L+ G +++

Sbjct: 159 VVAHLINYGTKGKVR 173

Score = 115 (17.3 bits), Expect = 1.8e-21, Sum P(2) = 1.8e-21
Identities = 37/119 (31%), Positives = 58/119 (48%)

Query: 497 ATPLHLACQKGYQSVTLTLLHYKASAEVQ--DNNGNTPLHLACTYGHEDCVKALVYYDVE 554
AT A + G ++ L H + ++ + NG LHLA GH V L++ ++
Sbjct: 13 ATSFLRAARG--NLDKALDHLRNGVDINTCNQNGNLGLHLASKEGHVKMVMVELLHKEII 70

Query: 555 SCRLDIGNEKGDTPHLHIAARWGYQGVETLLQNGASTEIQNRLKETPLKCALNSKILSVM 614
L+ +KG+T LHIAA G V+ L+ GA+ Q++ TPL A L V+
Sbjct: 71 ---LETTTKGNTALHIAALAGQDEVVRELNVYGANVNAQSQKGFPLYMAAQENHLEV 127

Query: 615 E 615
+
Sbjct: 128 K 128

Score = 106 (15.9 bits), Expect = 1.8e-01, Sum P(2) = 1.6e-01
Identities = 34/121 (28%), Positives = 54/121 (44%)

Query: 769 NAGARNADQAVPLHLACQGHFQVVKCLLDSNAKPNKKDLSGNTPLIYACSGGHHHELVAL 828
+ G R AD A A + G+ L + N + +G L A GH ++V
Sbjct: 4 SVGFREADAATSFRLAARSGNLDKALDHLRNGVDINTCNQNGNLGLHLASKEGHVKMVMVE 63

Query: 829 LLQHGAASINASNKGNTALHEAVIEKHVFFVVELLLHGASVQVLNKRQRTAVDCAEQNSK 888
LL + + KGNTALH A + VV L+ +GA+V +++ T + A Q +
Sbjct: 64 LLHKEIILETTTKGNTALHIAALAGQDEVVRELNVYGANVNAQSQKGFPLYMAAQENH 123

Query: 889 I 889
+
Sbjct: 124 L 124

Score = 40 (6.0 bits), Expect = 1.6e-14, Sum P(2) = 1.6e-14
Identities = 11/56 (19%), Positives = 23/56 (41%)

Query: 622 ERRQKSSEAPVQSPQRSVDSISQESSTSSFSMSAGSRQEETKKDYREVEKLLRAV 677
+RRQ+ E VQ + + + Q + + Q ++ +K++R V
Sbjct: 1614 DRRQQGQEEQVQEAKNTFTQVVQGNFQNIPEQVTEEQTDEQGNIVTKKIIRKV 1669

Score = 38 (5.7 bits), Expect = 2.6e-14, Sum P(2) = 2.6e-14
Identities = 6/12 (50%), Positives = 10/12 (83%)

Query: 806 KDLSGNTPLIYA 817
+D++G T L+YA
Sbjct: 1186 EDITGTTKLVYA 1197

Pedant information for DKFZphtes3_1817, frame 2

Report for DKFZphtes3_1817.2

[LENGTH] 1050
[MW] 117013.72
[pI] 6.47
[HOMOL] TREMBL:DMANKY_1 product: "ankyrin"; Drosophila melanogaster ankyrin mRNA,
complete cds. 2e-45
[FUNCAT] 08.19 cellular import [S. cerevisiae, YOR034c] 5e-13
[FUNCAT] 10.05.99 other pheromone response activities [S. cerevisiae, YDR264c]
3e-12
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins
[S. cerevisiae, YDR264c] 3e-12
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YIL112w] 2e-11
[FUNCAT] 06.13.01 cytoplasmic degradation [S. cerevisiae, YGR232w] 8e-10
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YIR033w] 2e-08
[FUNCAT] 04.05.01.07 chromatin modification [S. cerevisiae, YIR033w] 2e-08
[FUNCAT] 01.04.04 regulation of phosphate utilization [S. cerevisiae, YGR233c]
3e-08
[FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YML097c] 5e-05
[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YML097c]
5e-05
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YML097c] 5e-05
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YML097c]
5e-05
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YER111c] 3e-04
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YER111c] 3e-04
[BLOCKS] BL00901A Cysteine synthase/cystathionine beta-synthase P-phosphate att
[SCOP] dlawcb 1.91.3.1.2 GA binding protein (GABP) alpha GA bindini 4e-12
[EC] 3.1.3.53 Myosin-light-chain-phosphatase 1e-12
[PIRKW] phosphotransferase 1e-19
[PIRKW] nucleus 1e-13

[PIRKW] potassium channel 5e-15
 [PIRKW] early protein 2e-13
 [PIRKW] tumor suppressor 1e-09
 [PIRKW] duplication 1e-14
 [PIRKW] tandem repeat 1e-19
 [PIRKW] heterodimer 1e-14
 [PIRKW] potassium transport 5e-15
 [PIRKW] cell cycle control 1e-10
 [PIRKW] serine/threonine-specific protein kinase 1e-19
 [PIRKW] transmembrane protein 5e-15
 [PIRKW] transport protein 5e-15
 [PIRKW] DNA binding 2e-11
 [PIRKW] oncogene 1e-08
 [PIRKW] ATP 1e-19
 [PIRKW] protein kinase inhibitor 1e-09
 [PIRKW] voltage-gated ion channel 5e-15
 [PIRKW] phosphoprotein 4e-38
 [PIRKW] apoptosis 1e-19
 [PIRKW] liver 4e-09
 [PIRKW] integrin binding 3e-16
 [PIRKW] differentiation 2e-12
 [PIRKW] transforming protein 1e-08
 [PIRKW] alternative splicing 1e-40
 [PIRKW] coiled coil 1e-14
 [PIRKW] peripheral membrane protein 2e-38
 [PIRKW] transcription factor 4e-16
 [PIRKW] transcription regulation 2e-16
 [PIRKW] nucleotide binding 5e-15
 [PIRKW] phosphoric monoester hydrolase 1e-12
 [PIRKW] cytoskeleton 8e-39
 [PIRKW] calmodulin binding 1e-19
 [PIRKW] smooth muscle 1e-12
 [SUPFAM] ankyrin 1e-40
 [SUPFAM] death-associated protein kinase 1e-19
 [SUPFAM] ankyrin repeat homology 1e-40
 [SUPFAM] protein kinase homology 1e-19
 [SUPFAM] vaccinia virus 27.4K HindIII-C protein homology 3e-07
 [SUPFAM] int-3 transforming protein 1e-08
 [SUPFAM] unassigned ankyrin repeat proteins 2e-38
 [SUPFAM] notch protein 2e-12
 [SUPFAM] fowlpox virus BamHI-ORF7 protein 2e-13
 [SUPFAM] rel homology 2e-11
 [SUPFAM] EGF homology 2e-12
 [PROSITE] ATP_GTP_A 1
 [PFAM] Ank_repeat
 [KW] Irregular
 [KW] 3D
 [KW] LOW_COMPLEXITY 3.05 %

SEQ MALYDEDLLKNPFYLAQKCRPDLCSKVAQIHGIVLPCKGSLSSSIQSTCQFESYILIP
 SEG
 lawcB

 SEQ VEEHFQTLNGKDVFIQGNRIKLGAGFACLLSVPILFEETFYNEKEESFSILCIAHPLEKR
 SEG
 lawcB

 SEQ ESSEEPLAPSDPFSLKTIEDVREFLGRHSERFDRNIASFHRTFRECEKSLRHHIDSANA
 SEG
 lawcB

 SEQ LYTKCLQQLLRDLSHLKMLAKQEAQMNLKQAVEIYVHHEIYNLIFKYVGTMEASEDAAFN
 SEG
 lawcB

 SEQ KITRSLQDLQKQDIGVKPEFSFNIPRAKRELAQLNKCTSPQQLVCLRKVVQLITQSPSQ
 SEG
 lawcB

 SEQ RVNLETMCADDLLSVLLYLLVKTEIPNWMANLSYIKNFRFSSSLAKDELGYCLTSFEAAIE
 SEGxxxxxxxxx.....
 lawcB

 SEQ YIRQGSLSAKPPESEFGDRLFLKQRMSLLSQTSSPTDCLFKHIASGNQKEVERLLSQE
 SEG
 lawcB

 SEQ DHDKDTVQKMCHPLCFDCEKLVSGRLNDPSVVTPFSDRRGHTPLHVAAVCGQASLID
 SEG
 lawcB


```

SEQ    LLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKASAEVDNNGNTPLHLACTYG
SEG    .....
lawcB  .....

SEQ    HEDCVKALVYYDVESCRLDIGNEKGDTPHIAARWGYQGVIETLLQNGASTEIQNRLKET
SEG    .....
lawcB  .....

SEQ    PLKCALNSKILSVMEAYHLSFERRQKSSEAPVQSPQRSVDSISQESSTSSFSSMSAGSRQ
SEG    .....XXXXXXXXXXXXXXXXXXXXXXXXX
lawcB  .....

SEQ    EETKKDYREVEKLLRAVADGDLEMVRYLLEWTEEDLEDAEDTVSAADPEFCHPLCQCPKC
SEG    .....
lawcB  .....

SEQ    APAQKRLAKVPASGLGVNVTSDQGSSPLHVAALHGRADLIRLLKHGANAGARNADQAVP
SEG    .....
lawcB  .....CHHHHHHHHHHCHHHHHHHHHHCCCC-CCTTTTCCH

SEQ    LHLACQQGHFQVVKCLLDSNAKPNKKDLSGNTPLIYACSGGHHELVALLQHGASINASN
SEG    .....
lawcB  HHHHHHCHHHHHHHHHHCCCTTTTCTTTTCCHHHHHHHHTTHHHHHHHHHCCCTTTTEE

SEQ    NKGNTALHEAVIEKHVFVVELLLLHGASVQVLNKRQRTAVDCAEQNSKIMELLQVVPSCV
SEG    .....
lawcB  TTTEHHHHHHHHCHHHHHHHHHHCCCTTTTCBTTTBCHHHHHHHHCHHHHHHC.....

SEQ    ASLDDVAETDRKEYVTVKIRKKWNSKLYDLPDEPFTRQFYFVHSAGQFKGKTSREIMARD
SEG    .....
lawcB  .....

SEQ    RSVPNLTEGSLHEPGRQSVTLRQNNLPAQSGSHAAEKGNSDWPERPGLTQTGPGHRRMLR
SEG    .....
lawcB  .....

SEQ    RHTVEDAVVSQGPEAAGPLSTPQEVSASRS
SEG    .....
lawcB  .....

```

Prosites for DKFZphtes3_1817.2

PS00017 945->953 ATP_GTP_A PDOC00017

Pfam for DKFZphtes3_1817.2

```

HMM_NAME      Ank repeat
HMM            *GyTPLHIAARYNNvEMVrLLQHGADIN*
               G+TPLH+AA  ++  ++++LL+++GA +N
Query          463  GHTPLHVAAVCGQASLIDLLVSKGAMVN      490

32.12 (bits) f: 496 t: 523 Target: dkfzphes3_1817.2 similarity to ankyrins
Alignment to HMM consensus:
Query          *GyTPLHIAARYNNvEMVrLLQHGADIN*
               G TPLH+A++ +  ++  LLL + A+
dkfzphes3      496  GATPLHLACQKGYQSVTLLLLHYKASAE      523

Query          f: 529 t: 556 Target: dkfzphes3_1817.2 similarity to ankyrins
Alignment to HMM consensus:
HMM            *GyTPLHIAARYNNvEMVrLLQHGADIN*
               G+TPLH+A+ Y+++++V+ L+ +
Query          529  GNTPLHLACTYGHEDCVKALVYYDVESC      556

42.65 (bits) f: 565 t: 592 Target: dkfzphes3_1817.2 similarity to ankyrins
Alignment to HMM consensus:
Query          *GyTPLHIAARYNNvEMVrLLQHGADIN*
               G+TPLHIAAR +  +++ LLQ+GA+
dkfzphes3      565  GDTPLHIAARWGYQGVIETLLQNGASTE      592

Query          f: 744 t: 771 Target: dkfzphes3_1817.2 similarity to ankyrins
Alignment to HMM consensus:
HMM            *GyTPLHIAARYNNvEMVrLLQHGADIN*
               G +PLH+AA  +++  +++RLLL+HGA+
Query          744  GSSPLHVAALHGRADLIRLLLLKHGANAG      771

```

36.38 (bits) f: 777 t: 804 Target: dkfzphes3_1817.2 similarity to ankyrins

Alignment to HMM consensus:

Query *GyTPLHIAARyNNvEMVrLLQHGA⁺DIN*
PLH+A++++ ++V+ LL+ +A +N
dkfzphes3 777 QAVPLHLACQGGHFQVVKCLLDSNAKPN 804

Query f: 810 t: 837 Target: dkfzphes3_1817.2 similarity to ankyrins

Alignment to HMM consensus:

HMM *GyTPLHIAARyNNvEMVrLLQHGA⁺DIN*
G+TPL++A+ ++ E+V LLLQHGA+IN
Query 810 GNTPLIYACSGGHHELVALLQHGA⁺SIN 837

44.62 (bits) f: 843 t: 870 Target: dkfzphes3_1817.2 similarity to ankyrins

Alignment to HMM consensus:

Query *GyTPLHIAARyNNvEMVrLLQHGA⁺DIN*
G+T+LH A+++ +V +V+LLL HGA++
dkfzphes3 843 GNTALHEAVIEKHVFVVELLLLHGASVQ 870

DKFZphtes3_19f19

group: testes derived

DKFZphtes3_19f19 encodes a novel 254 amino acid protein with weak similarity to *S. cerevisiae* protein YFL046w.

The protein contains a RGD cell attachment site.
No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to YFL046w

localisation: 3 STS match perfect but HS1292427 matches to chromosome 4

Sequenced by MediGenomix

Locus: /map="405.0/.3 cR from top of Chr11 linkage group"

Insert length: 1395 bp
Poly A stretch at pos. 1367, no polyadenylation signal found

```
1 GGGACCACGG TGGCGCCTGC GCTGGGAGGT GAGCTTGTGA CAGAGCGAAA
51 ACTACAATTC CCAGCATTCC TGTGGTGCCA GAACTACCTT GCCCGAAAAGC
101 CTGTGCGAGA TTTACCCCGT CTTCCGCCTC CCTCCACCG GAAACTCTG
151 AGGACATGAA TAGTCGCCAG GCTTGGCGGC TCTTTCTCTC CCAAGGCAGA
201 GGAGATCGTT GGGTTTCAAG GCCCGCGGG CATTCTCGC CGGCCCTGCG
251 GAGAGAGTTC TTTACTACCA CAACCAAGGA GGGATATGAT AGGCGGCCAG
301 TGGATATAAC TCCTTTAGAA CAAAGGAAAT TAACTTTTGA TACCCATGCA
351 TTGGTTCAGG ACTTGGAAC TCATGGATTT GACAAAACAC AAGCAGAAAC
401 AATTGTATCA GCGTTAACTG CTTTATCAAA TGTGAGCTG GATACTATCT
451 ATAAAGAGAT GGTCACCTCA GCTCAACAGG AAATAACAGT ACAACAGCTA
501 ATGGCTCATT TGGATGCTAT CAGGAAAGAC ATGGTCATCC TAGAGAAAAG
551 TGAATTGCA AATCTGAGAG CAGAGAATGA GAAATGAAA ATTGAATTAG
601 ACCAAGTTAA GCAACAACCTA ATGCATGAAA CCAGTCGAAT CAGAGCAGAT
651 AATAAACTGG ATATCAACTT AGAAAGGAGC AGAGTAACAG ATATGTTTAC
701 AGATCAAGAA AAGCAACTTA TGGAAACAAC TACAGAATTT AAAAAAAGG
751 ATACTCAAAC CAAAAGTATT ATTTAGAGA CCAGTAATAA AATTGACGCT
801 GAAATTGCTT CCTTAAAAAC ACTGATGGAA TCTAACAAAC TTGAGACAAT
851 TCGTTATCTT GCAGCTTCGG TGTCTACTTG CCTGGCAATA GCATTGGGAT
901 TTTATAGATT CTGGAAGTAG TATTAATGCT CATCTGCTG TGGCTGTGG
951 CTCTTAGAAA CACCAACCG GGAGAGATTT ACTTTGAACA TTGTCAGTTG
1001 CAGCAAAAAT TTTACTACCA AGATTATTCG AAGTGATATC GGACTAAAAG
1051 AGGAAGTGTT TTAGAATGAG AAGAGATACT GTGTCTTTAT TGTGTGTGTG
1101 TGAGTGCAGG TGTGTGTCTT TATTATATTG AAAAGCTGTC ACTCAGACCT
1151 GGTTTGAGAT AGAAGAGCAT TTTGTCCTTT TGATAGTTAA TAGAAATTGA
1201 ACCAGAGTTT TCTTATGTTT GCTTGAACAG TTGTGTAAAT CATACAGGAT
1251 TTTGTGGGTA TTGGTTGAAT ATTTGTAAAC CATTCCTAG CCTACATATT
1301 TATTACTGAA TTAACCTTCC TGATAACCAT TGCATATTA CATTTTCTA
1351 TAAATGAAA GATTATTACA AAAAAA AAAA
```

BLAST Results

Entry HS419346 from database EMBL:
human STS WI-13569.
Score = 2154, P = 8.6e-91, identities = 446/459

Entry HS1292427 from database EMBL:
human STS SHGC-50338.
Score = 1737, P = 7.2e-72, identities = 359/369

Entry HS253344 from database EMBL:
human STS WI-13893.
Score = 1578, P = 1.0e-64, identities = 358/397

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 156 bp to 917 bp; peptide length: 254
 Category: similarity to unknown protein
 Classification: no clue
 Prosite motifs: RGD (15-18)

```

1 MNSRQAWRLF LSQGRGDRWV SRPRGHFSPA LRREFFTTTT KEGYDRRPVD
51 ITPLEQRKLT FDTHALVQDL ETHGFDKTQA ETIVSALTAL SNVSLDTIYK
101 EMVTQAQQEI TVQQLMAHLD AIRKDMVILE KSEFANLRAE NEKMKIELDQ
151 VKQQLMHETS RIRADNKLDI NLERSRVTDM FTDQEKQLME TTTEFTKKDT
201 QTKSISETS NKIDAEIASL KTLMESNKLE TIRYLAASVF TCLAIALGFY
251 RFWK

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_19f19, frame 3

SWISSPROT:YAN8_SCHPO HYPOTHETICAL 24.6 KD PROTEIN C3H1.08 IN CHROMOSOME I., N = 1, Score = 144, P = 8.4e-09

PIR:S56209 probable membrane protein YFL046w - yeast (*Saccharomyces cerevisiae*), N = 1, Score = 138, P = 5.4e-08

>SWISSPROT:YAN8_SCHPO HYPOTHETICAL 24.6 KD PROTEIN C3H1.08 IN CHROMOSOME I.
 Length = 211

HSPs:

Score = 144 (21.6 bits), Expect = 8.4e-09, P = 8.4e-09
 Identities = 34/121 (28%), Positives = 67/121 (55%)

```

Query:   70 LETHGFDKTQAETIVSALTALSNVSLDTIYKEMVTQAQQE-ITVQQLMAHLDIAIRKDMVI 128
          LE  G+  AETI + +  ++ +L + K + +A+QE ++ QQ  L  IRK +
Sbjct:   46 LEQAGYSVKNAETITNLMRTITGEALTELEKNIGFKAKQESVSFQQKRTFLQ-IRKYLET 104

Query:   129 LEKSEFANLRAENKMKIELDQVKQQLMHETSRIADNKLDINLERSRVTDMFTDQEKQL 188
          +E++EF +R  ++K+  E+++ K  L  +  ++ +L++NLE+ R+ D  T  +  +
Sbjct:   105 IEENEFDKVRKSSDKLINEIEKTKSSLREDVKTALSEVRLNLEKGRMKDAATSRNTNI 164

Query:   189 ME 190
          E
Sbjct:   165 HE 166

```

Pedant information for DKFZphtes3_19f19, frame 3

Report for DKFZphtes3_19f19.3

```

{LENGTH}      254
{MW}           29505.73
{pI}           6.99
{HOMOL}        PIR:S56209 probable membrane protein YFL046w - yeast (Saccharomyces cerevisiae)
2e-10
{FUNCAT}       99 unclassified proteins      [S. cerevisiae, YFL046w] 8e-12
{PROSITE}      RGD 1
{KW}           TRANSMEMBRANE 1
{KW}           LOW_COMPLEXITY 5.12 %
{KW}           COILED_COIL 11.02 %

```

```

SEQ  MNSRQAWRLFLSQGRGDRWVSRPRGHFSPALRREFFTTTTKEGYDRRPVDITPLEQRKLT
SEG  .....
PRD  ccchhhhhhhhhccccceeeccccchhhhhhhheeeccccccccccchhhhhhhcc
COILS .....
MEM  .....

```

```

SEQ  FDTHALVQDLETHGFDKTQAETIVSALTALSNVSLDTIYKEMVTQAQQEITVQQLMAHLD
SEG  .....
PRD  chhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....

```

Prosites for DKFZphtes3 19f19.3

(No Pfam data available for DKFZphtes3 19f19.3)

DKFZphtes3_19j17

group: testes derived

DKFZphtes3_19j17 encodes a novel 436 amino acid protein with partial similarity to C.elegans Y40B1A.2 protein.

The novel protein contains two Prosite WW/rsp5/WWP domain signatures. The WW domain (or rsp5 or WWP domain) has been originally discovered as a short conserved region in a number of unrelated proteins, such as dystrophin, utrophin, vertebrate YAP protein, mouse NEDD-4 and yeast RSP5. The domain is repeated up to 4 times in some proteins. It has been shown to bind proteins with particular proline-motifs, [AP]-P-P-[AP]-Y, and thus resembles somewhat SH3 domains. It appears to contain beta-strands grouped around four conserved aromatic positions; generally Trp. The name WW or WWP derives from the presence of these Trp as well as that of a conserved Pro. It is frequently associated with other domains typical for proteins in signal transduction processes.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to C.elegans Y40B1A.2

there are two long ORFs in this cDNA according to EST:
HS12146/HS75086/AA923755/MMAA17335 remaining intron at Bp 1506-1733

Sequenced by MediGenomix

Locus: unknown

Insert length: 2762 bp
Poly A stretch at pos. 2740, no polyadenylation signal found

```
1 ATTCTCAGCC AAATTTTTTT ATTTTGTGCA GAATCAGTGT GCAAGGTGGT
51 TTATAAGATA ATGGAGTGGT TTTTGTGTG GTTAGTGTG ATTGTTATC
101 AGGAGTCTTA TTGTAACGCT TAAGCATTAG GTTTTGTGTC TGAGAACTT
151 TAAAGAGTAA AGCAGAATTG AAAGTGGAAA TTTTAATTTT GTAAGTTCAT
201 AAAATTTAAT GATAATACAC CAAAGTTTAT GTTTAAATTA GGGAGTTTAA
251 GGTTCGAATT CTTTCTCTTT TTTTGTGGG GGGTGATGTT TTACAGGCAC
301 TTAAGTATTC ATCGAAGAGT CACCCAGTA GCGGTGATCA CAGACATGAA
351 AAGATGCGAG ACGCCGGAGA TCCTTCACCA CCAAATAAAA TGTGCGGAG
401 ATCTGATAGT CCGTGAACA AATACAGTGA CAGCACAGGT CACAGTAAGG
451 CCAAAAATGT GCATACTCAC AGAGTTAGAG AGAGGGATGG TGGGACCACT
501 TACTCTCCAC AAGAAAATTC ACACAACCAC AGTGCTCTTC ATAGTTCAAA
551 TTCACATTCT TCTAATCCAA GCAATAACCC AAGCAAAACT TCAGATGCAC
601 CTTATGATTG TGCAGATGAC TGGTCTGAGC ATATTAGCTC TTCTGGGAAA
651 AAGTACTACT ACAATTGTCG AACAGAAGTT TCACAATGGG AAAAACCAAA
701 AGAGTGGCTT GAAAGAGAAC AGAGACAAA AGAAGCAAA AGATGGCAG
751 TCAACAGCTT CCCAAAAGAT AGGGATTACA GAAGAGAGGT GATGCAAGCA
801 ACAGCCACTA GTGGGTTTGC CAGTGAATG GAAGACAAGC ATTCCAGTGA
851 TGCCAGTAGT TTGCTCCAC AGAATATTTT GTCTCAACA AGCAGACACA
901 ATGACAGAGA CTACAGACTG CCAAGAGCAG AGACTCACAG TAGTTCTACG
951 CCAGTACAGC ACCCATCAA ACCAGTGGT CATCCAAGT CTACCCCAAG
1001 CACTGTTCCT TCTAGTCCAT TTACGCTACA GTCTGATCAC CAGCCAAAAGA
1051 AATCATTTGA TGCTAATGGA GCATCTACTT TATCAAACT GCCTACACCC
1101 ACATCTTCTG TCCCTGCACA GAAACAGAA AGAAAAGAA CTACATCAGG
1151 AGACAAACCC GTATCACATT CTGACACAAC TCCTTCCACG TCTTCTGCCT
1201 CTGGACTGAA CCCCATCTCT GCACCTCCAA CATCTGCTTC AGCGGTCCCT
1251 GTTCTCTCTG TTCCACAGTC GCCAATACCT CCCTTACTTC AGGACCCAAA
1301 TCTTCTTAGA CAATTGCTTC CTGCTTTGCA AGCCACGCTG CAGCTTAATA
1351 ATTCTAATGT GGACATATCT AAAATAAATG AAGTTCTTAC AGCAGCTGTG
1401 ACACAAGCCT CACTGCAGTC TATAATTCAT AAGTTTCTTA CTGCTGGACC
1451 ATCTGCTTTC AACATAACGT CTCTGATTTT TCAAGCTGCT CAGCTCTCTA
1501 CACAAGATAT CCCCTCTCAT GAAGGTATCC AAATGGAGAG AGATACACAT
1551 AGGAGCAATG GGGAAAGTGA AGGGTCACTT TGTCAGAAAG CTGATAAACA
1601 GCAGGAATGC CTTGTCTGGA ATGGAAGTAT AATGGTGCAA AGACTCTTGC
1651 AACCTCTTGG CTAGCCTCAT GAGCAGGAGA CTGCGTGGGA TACCTGGGCC
1701 TAAATGTAGA ATAAGAAAGA AGAAATAAGG ATGCCAGCC ATCTAATCAG
1751 TCTCCGATGT CTTTAACATC TGATGCGTCA TCCCAAGAT CATATGTTTC
1801 TCCAAGAATA AGCACACCTC AAACTAACAC AGTCCCTATC AAACCTTTGA
1851 TCAGTACTCC TCCTGTTTCA TCACAGCCAA AGGTTAGTAC TCCAGTAGTT
1901 AAGCAAGGAC CAGTGTACA GTCAGCCACA CAGCAGCCTG TAACTGCTGA
1951 CAAGCAGCAA GGTCTGAAC CTGTCTCTCC TCGAAGTCTT CAGCGCTCAA
2001 GCCAGAGAAG TCCATCACCT GGTCCCAATC ATACTTCTAA TAGTAGTAAT
2051 GCATCAATG CAACAGTTGT ACCACAGAAT TCTTCTGCCC GATCCACGTG
```

```

2101 TTCATTAACG CCTGCACTAG CAGCACACTT CAGTGAAAAT CTCATAAAAC
2151 ACGTTCAAGG ATGGCCTGCA GATCATGCAG AGAAGCAGGC ATCAAGATTA
2201 CGCGAAGAAG CGCATAACAT GGGAACTATT CACATGTCCG AAATTTGTAC
2251 TGAATTAAAA AATTTAAGAT CTTAGTCCG AGTATGTGAA ATTCAAGCAA
2301 CTTTGGGAGA GCAAAGGATA CTATTTTGA GACAACAAAT TAAGGAACCT
2351 GAAAAGCTAA AAAATCAGAA TTCCTTCATG GTGTGAAGAT GTGAATAATT
2401 GCACATGGTT TTGAGAACAG GAACTGTAAA TCTGTTGCC AATCTTAACA
2451 TTTTGTAGCT GCATTTAAGT AGACTTTGGA CCGTTAAGCT GGGCAAAGGA
2501 AATGACAAGG GGACGGGGTC TGTGAGAGTC AATTCAGGGG AAAGATACAA
2551 GATTGATTG TAAAACCCTT GAAATGTAGA TTTCTTGTAG ATGTATCCTT
2601 CACGTTGTAA ATATGTTTGT TAGAGTGAAG CCATGGGAAG CCATGTGTAA
2651 CAGAGCTTAG ACATCCAAAA CTAATCAATG CTGAGGTGGC TAAATACCTA
2701 GCCTTTTACA TGTAAACCTG TCTGCAAAAT TAGCTTTTTT AAAAAAAAAA
2751 AAAAAAAAAA AA

```

BLAST Results

Entry AC005876 from database EMBLNEW:
Homo sapiens chromosome 10 clone CIT987SK-1188I5 map 10p11.2-10p12.1,
complete sequence.
Score = 2130, P = 0.0e+00, identities = 426/426
12 exons matching Bp 492-2740

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 1757 bp to 2383 bp; peptide length: 209
Category: questionable ORF
Classification: no clue

```

1 MSLTSDASSP RSYVSPRIST PQNTVPPIKP LISTPPVSSQ PKVSTPVVKQ
51 GPVSQSATQQ PVTADKQQGH EPVSPRSLQR SSQRSPSPGP NHTSNSSNAS
101 NATVVPQNSS ARSTCSLTPA LAAHFSENLI KHVQGWPAKH AEKQASRLRE
151 EAHNMGTIHM SEICTELKNL RSLVRVCEIQ ATLREQRILF LRQIQI KELEK
201 LKNQNSFMV

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_19j17, frame 2

No Alert BLASTP hits found

Peptide information for frame 3

ORF from 354 bp to 1661 bp; peptide length: 436
Category: similarity to unknown protein
Classification: unclassified
Prosites motifs: WW DOMAIN_1 (90-116)
WW_DOMAIN_1 (90-116)

```

1 MRDAGDPSP NKMLRRSDSP ENKYSDSTGH SKAKNVHTR VRERDGGTSY
51 SPQENSHNHS ALHSSNSHSS NPSNNPSKTS DAPYDSADDW SEHISSSGKK
101 YYNCRTEVS QWEKPKEWLE REQRQKEANK MAVNSFPKDR DYRREVMQAT
151 ATSGFASGME DKHSSDASSL LPQNILSQTs RHNDRDYRLP RAETHSSSTP
201 VQHPIKPVVH PTATPSTVPS SPFTLQSDHQ PKKSF DANGA STLKSLPTPT
251 SSVPAQKTER KESTSGDKPV SHSCTTPSTS SASGLNPTSA PPTSASAVPV
301 SPVPQSPIPP LLQDPNLLRQ LLPALQATLQ LNNSNVDISK INEVLTAAVT
351 QASLQSIHK FLTAGPSAFN ITSLISQAAQ LSTQDIPLHE GIQMERDTHR
401 SKWEVKGSLC QKADKQCECL VWNGSIMVQR LLQPSG

```

BLASTP hits

656


```

SEQ  MRDAGDPSPPNKMRLRRSDSPENKYSdstGHSKAKNVHthrvrERDGGTSYSPQENSHNHS
SEG  .....xxxxxx
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  ALHSSNSHSSNPSNNPSKTSdAPYDSADDWSEHISSSGKKYYNCRTEVSQWEKPKEWLE
SEG  .....xxxxxx
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccchhhh

SEQ  REQRQKEANKMAVNSFPKDRDYRREVMQATATSGFASGMEDKHSSDASSLLPQNILSQTs
SEG  .....
PRD  hhhhhhhhhhhhhccccccccchhhhhhhhhhhcccccccccccccccccccccccccccccccccc

SEQ  RHNDRDYRLPRAETHSSSTPVQHPIKPVVHPTATPSTVPSSPFTLQSDHQPKKsFDANGA
SEG  .....xxxxxx
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  STLSKLPTPTSSVPAQKTERKESTSGDKPVSHSCTTPSTSSASGLNPTSAPPTSASAVPV
SEG  .....xxxxxx
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  SPVPQSPiPPLLDQPNLLRQLLPALQATLQLNNSNVDISKINEVLTAAVTQASLQSIHK
SEG  .....xxxxxx
PRD  cccccccccccccccccchhhhhhhhhhhhhccccccccccccchhhhhhhhhhhhhhhhhhhhh

SEQ  FLTAGPSAFNITSLISQAQLSTQDIPLHEGIQMERDTHRSKWEVKGSLCQKADKQOECL
SEG  .....
PRD  hhccccccccchhhhhhhhhhhhhccccccccccccccccccccccccccccchhhhhhhccce

SEQ  VWNQSIMVQRLLPQPSG
SEG  .....
PRD  eccccchhhhhhhcccccc

```

Prosite for DKFZphtes3_19j17.3

```

PS01159  90->116    WW_DOMAIN_1      PDOC50020
PS01159  90->116    WW_DOMAIN_1      PDOC50020

```

Pfam for DKFZphtes3_19j17.3

HMM_NAME WW/rsp5/WWP domain containing proteins

```

HMM      *LPsGWEeHWDpsGRpWYYWNHETkTTQWEpP*
+ ++W EH++ SG+ YY+N T+ +QWE+P
Query    86 SADDWSEHISSSGKK-YYNCRTEVSQWEKP 115

```

DKFZphtes3_lcl

group: signal transduction

DKFZphtes3_lcl encodes a novel 632 amino acid putative GTPase-activating protein, related to drosophila rotund transcript and human n-chimaerin.

rac small GTPase is associated with type-I phosphatidylinositol 4-phosphate 5-kinase and regulating the production of phosphatidylinositol 4,5-bisphosphate. The new protein is expected to activate p21rac-related small GTPases.

The new protein can find application in modulating/blocking the response to a cellular receptor.

similarity to GTPase-activating proteins

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 3237 bp

Poly A stretch at pos. 3227, no polyadenylation signal found

```

1  GCGAAGTGAA GGGTGGGCCA GGTGGGGCCA GGCTGACTGA ATGTATCTCC
51 TAGCTATGGA CTAATAATA CATGGGGGGA AATAACAAG TATTATGAG
101 GGTGAAAATG TGACCCAGCA GGAAAATTAC AACTATTTC AATTGACGTT
151 GAATAGGATG AGTCATGGAA TTTAAGTGAT TTAAGAAGA TTATACTACT
201 GGTAGATAGA AGAGCTAAAG AAAGATGGAT ACTATGATGC TGAATGTGCG
251 GAATCTGTTT GAGCAGCTTG TGGCCCGGGT GGAGATTCTC AGTGAAGGAA
301 ATGAAGTGCA ATTTATCCAG TTGGCGAAGG ACTTTGAGGA TTTCCTGAAA
351 AAGTGGCAGA GGAAGTACCA TGAGCTGGGG AAATACAAGG ATCTTTTGAT
401 GAAAGCAGAG ACTGAGCGAA GTGCTCTGGA TGTAAAGCTG AAGCATGCAC
451 GTAATCAGGT GGATGTAGAG ATCAAACGGA GACAGAGAGC TGAGGCTGAC
501 TGGGAAAAGC TGGAAACGACA GATTACAGCTG ATTCGAGAGA TGCTCATGTG
551 TGACACATCT GGCAGCATTG AACTAAGCGA GGAGCAAAA TCAGCTCTGG
601 CTTTTCTCAA CAGAGGCCAA CCATCCAGCA GCAATGCTGG GAACAAAAGA
651 CTATCAACCA TTGATGAATC TGGTTCATT TTATCAGATA TCAGCTTTGA
701 CAAGACTGAT GAATCACTGG ATTGGGACTC TTCTTTGGTG AAGACTTTCA
751 AACTGAAGAA GAGAGAAAAG AGGCGCTCTA CTAGCCGACA GTTTGTTGAT
801 GGTCCCCCTG GACCTGTAAA GAAAACCTCG TCCATTGGCT CTGCAGTAGA
851 CCAGGGGGAAT GAATCCATAG TTGCAAAAAC TACAGTGACT GTTCCCAATG
901 ATGCGGGGCC CATCGAAGCT GTGTCCACTA TTGAGACTGT GCCATATTGG
951 ACCAGGAGCC GAAGGAAAAC AGGTACTTTA CAACCTTGGG ACAGTGACTC
1001 CACCCCTGAAC AGCAGGCAGC TGGAGCCAAG AACTGAGACA GACAGTGTGG
1051 GCACGCCACA GAGTAATGGA GGGATGCGCC TGCATGACTT TGTTTCTAAG
1101 ACGGTTATTA AACCTGAATC CTGTGTTCCA TGTGGAAGC GGATAAAAT
1151 TGGCAAAATTA TCTCTGAAGT GTGAGAGCTG TCGTGTGGTC TCTCATCCAG
1201 AATGTGCGGA CCGCTGTCCC CTTCCCTGCA TTCCTACCCT GATAGGAACA
1251 CCTGTCAAGA TTGGAGAGGG AATGCTGGCA GACTTTGTGT CCCAGACTTC
1301 TCCAATGATC CCCTCCATTG TTGTGCATTG TGTAATGAG ATTGAGCAAA
1351 GAGGTCTGAC TGAGACAGGC CTGTATAGGA TCTCTGGCTG TGACCCGACA
1401 GTAAAAGAGC TGAAGAGAA ATTCCCTCAG GTGAAAAGTG TACCCCTCCT
1451 CAGCAAAGTG GATGATATCC ATGCTATCTG TAGCCTTCTA AAAGACTTTC
1501 TTCGAAACCT CAAAGAACCT CTTCTGACCT TTCGCCTTAA CAGAGCCTTT
1551 ATGGAAGCAG CAGAAATCAC AGATGAAGAC AACAGCATAG CTGCCATGTA
1601 CCAAGCTGTT GGTGAAGTGC CCCAGGCCAA CAGGACACA TTAGCTTTCC
1651 TCATGATTCA CTTGCAGAGA GTGGCTCAGA GTCCACATAC TAAATGGAT
1701 GTTGCCAATC TGGCTAAAGT CTTTGGCCCT ACAATAGTGG CCCATGCTGT
1751 GCCCAATCCA GACCCAGTGA CAATGTTACA GGACATCAAG CGTCAACCCA
1801 AGGTGGTTGA GCGCCTGCTT TCCTTGCTCT TGGAGTATTG GAGTCAGTTC
1851 ATGATGGTGG AGCAAGAGAA CATTGACCCC CTACATGTCA TTGAAAACTC
1901 AAATGCCTTT TCAACACCAC AGACACCAGA TATTAAAGTG AGTTTACTGG
1951 GACCTGTGAC CACTCTGAA CATCAGCTTC TCAAGACTCC TTCATCTAGT
2001 TCCCTGTGAC AGAGAGTCCG TTCCACCCTC ACCAAGAACA CTCTAGATT
2051 TGGGAGCAAA AGCAAGTCTG CCACTAACCT AGGACGACAA GGCACACTTT
2101 TTGCTTCTCC AATGCTCAAG TGAAGTCACA TCTGCCTGTT ACTTCCAGC
2151 ATTGACTGAC TATAAGAAA GACACATCTG TACTCTGCTC TGCAGCCTCC
2201 TGTAATCATT ACTACTTTTA GCATTCTCCA GGCTTTTACT CAAGTTTAAT
2251 TGTGCATGAG GGTTTTATTA AAACATATA TATCTCCCTT TCCTTCTCCT
2301 CAAGTCACAT AATATCAGCA CTTTGTGCTG GTCATTGTTG GGAGCTTTTA
2351 GATGAGACAT CTTTCCAGGG GTAGAAGGGT TAGTATGGAA TTGGTTGTGA
2401 TTTCTTTTGG GGAAGGGGGT TATTGTTCTT TTGGCTTAAA GCCAAATGCT
2451 GCTCATAGAA TGATCTTCTT CTAGTTTCAT TTAGAAGTGA TTTCCGTGAG
2501 ACAATGACAG AAACCTTACC TATCTGATAA GATTAGCTTG TCTCAGGGTG
2551 GGAAGTGGGA GGGCAGGGCA AAGAAAGGAT TAGACCAGAG GATTAGGAT

```

```

2601 GCCTCCTTCT AAGAACCAGA AGTTCTCATT CCCCATATG AACTGAGCTA
2651 TAATATGGAG CTTTCATAAA AATGGGATGC ATTGAGGACA GAACTAGTGA
2701 TGGGAGTATG CGTAGCTTTG ATTTGGATGA TTAGGTCTTT AATAGTGTG
2751 AGTGGCACAA CCTTGTAAT GTGAAAGTAC AACTCGTATT TATCTCTGAT
2801 GTGCCGCTGG CTGAACCTTG GGTTCATTTG GGGTCAAAGC CAGTTTTTCT
2851 TTAAATTTG AATTCATTCT GATGCTTGGC CCCCATACCC CCAACCTTGT
2901 CCAGTGGAGC CCAACTTCTA AAGGTCAATA TATCATCCTT TGGCATCCCA
2951 ACTAACATA AAGAGTAGGC TATAAGGGAA GATTGTCAAT ATTTTGTGGT
3001 AAGAAAAGCT ACAGTCATT TTTCTTTGCA CTTGGATGC TGAAATTTT
3051 CCCATGGAAC ATAGCCACAT CTAGATAGAT GTGAGCTTTT TCTTCTGTTA
3101 AAATTATTCT TAATGTCTGT AAAAACGATT TTCTTCTGTA GAATGTTTGA
3151 CTTCGTATTG ACCCTTATCT GTAAACACC TATTTGGGAT AATATTGGA
3201 AAAAAAGTAA ATAGCTTTTT CAAAATGAAA AAAAAA

```

BLAST Results

Entry U82984 from database EMBLEST:
Homo sapiens DRES 56 mRNA sequence.
Score = 8775, P = 0.0e+00, identities = 1757/1758
matches 3' end

Medline entries

93074974:
Developmental regulation and neuronal expression of the mRNA of rat
n-chimaerin, a
p21rac GAP:cDNA sequence.

93024458:
A Drosophila rotund transcript expressed during spermatogenesis and
imaginal disc
morphogenesis encodes a protein which is similar to human Rac
GTPase-activating
(racGAP) proteins.

Peptide information for frame 3

ORF from 225 bp to 2120 bp; peptide length: 632
Category: similarity to known protein

```

1 MDTMMLNVRN LFEQLVRRVE ILSEGNEVQF IQLAKDFEDF RKKWQRTDHE
51 LKGYKDLLMK AETERSALDV KLKHARNQVD VEIKRRQRAE ADCEKLERQI
101 QLIREMLMCD TSGSIQLSEE QKSALAFLNR GQPSSSNAGN KRLSTIDESG
151 SILSDISFDK TDESLDWDSS LVKTFKLKRR EKRRSTSRQF VDGPPGPVKK
201 TRSIGSAVDQ GNESIVAKTT VVVPNDGGPI EAVSTIETVP YWTRSRRTG
251 TLQPWNSDST LNSRQLEPRT ETDSVGTQPS NGGMRHLHDFV SKTVIKPESC
301 VPCGKRIKFG KLSLKRDCR VVSHPECRDR CPLPCIPTLI GTPVKIGEGM
351 LADFVSQTSF MIPSIIVHCV NEIEQRLGTE TGLYRISGCD RTVKELKEKF
401 LRVKTVPLLS KVDDIHAICS LLKDFLRNLK EPLLTFRNLN AFMEAAEITD
451 EDNSIAAMYQ AVGELPQANR DTLAFLMIHL QRVAQSPHTK MDVANLAKVF
501 GPTIVAHAVP NPDPVTMLQD IKRQPKVVER LLSLPLEYWS QFMMVEQENI
551 DPLHVIENSF AFSTPQTPDI KVSLLGPVTT PEHQLLKTPS SSSLSQRVRS
601 TLTKNTPRFG SKSKSATNLG RQGNFFASPM LK

```

BLASTP hits

Entry CEK08E3_4 from database TREMBLNEW:
gene: "K08E3.6"; Caenorhabditis elegans cosmid K08E3
Score = 452, P = 2.6e-48, identities = 126/377, positives = 189/377

Entry A48122 from database PIR:
GTPase-activating protein Rac homolog, splice form clone pcl.7 - fruit
fly (Drosophila melanogaster) (fragment)
Score = 480, P = 9.2e-46, identities = 111/270, positives = 155/270

Entry B48122 from database PIR:
GTPase-activating protein Rac homolog, splice form clone pcl.7d - fruit
fly (Drosophila melanogaster)
Score = 480, P = 9.2e-46, identities = 111/270, positives = 155/270

Entry DM22539_1 from database TREMBL:
 gene: "rotund"; product: "rnracGAP"; Drosophila melanogaster rnracGAP
 (rotund) gene, complete cds.
 Score = 480, P = 9.2e-46, identities = 111/270, positives = 155/270

Entry S29128 from database PIR:
 N-chimerin - rat
 Score = 336, P = 8.8e-30, identities = 86/253, positives = 128/253

Alert BLASTP hits for DKFZphtes3_lcl, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_lcl, frame 3

Report for DKFZphtes3_lcl.3

[LENGTH] 632
 [MW] 71026.84
 [pI] 9.08
 [HOMOL] PIR:B48122 GTPase-activating protein Rac homolog, splice form clone pcl.7d -
 fruit fly (Drosophila melanogaster) 2e-46
 [FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YBR260c] 3e-12
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YER155c] 2e-11
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YER155c] 2e-11
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YER155c]
 2e-11
 [FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YDL240w] 3e-09
 [FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YOR134w] 4e-09
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YOR134w] 4e-09
 [FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins
 [S. cerevisiae, YOR127w] 5e-09
 [FUNCAT] 09.04 biogenesis of cytoskeleton [S. cerevisiae, YPL115c] 3e-08
 [FUNCAT] 10.02.09 regulation of g-protein activity [S. cerevisiae, YPL115c] 3e-08
 [BLOCKS] BL00479B Phorbol esters / diacylglycerol binding domain proteins
 [BLOCKS] BL00479A Phorbol esters / diacylglycerol binding domain proteins
 [SCOP] dlpbwa_1.83.1.1.2 p85 alpha subunit RhoGAP domain [human (Homo sapiens)] 1e-55
 [SCOP] dirgp_1.83.1.1.1 p50 RhoGAP domain [human (Homo sapiens)] 1e-49
 [PIRKW] breakpoint cluster region 1e-19
 [PIRKW] transmembrane protein 7e-08
 [PIRKW] brain 3e-22
 [PIRKW] alternative splicing 1e-19
 [PIRKW] P-loop 2e-25
 [SUPFAM] CDC24 homology 3e-22
 [SUPFAM] bcr protein 3e-22
 [SUPFAM] myosin motor domain homology 2e-25
 [SUPFAM] pleckstrin repeat homology 4e-10
 [SUPFAM] LIM metal-binding repeat homology 2e-09
 [SUPFAM] protein kinase C zinc-binding repeat homology 5e-29
 [PROSITE] MYRISTYL 6
 [PROSITE] AMIDATION 1
 [PROSITE] CAMP_PHOSPHO_SITE 3
 [PROSITE] CK2_PHOSPHO_SITE 13
 [PROSITE] TYR_PHOSPHO_SITE 2
 [PROSITE] PKC_PHOSPHO_SITE 9
 [PROSITE] ASN_GLYCOSYLATION 1
 [PROSITE] DAG_PE_BINDING_DOMAIN 1
 [PFAM] Phorbol esters / diacylglycerol binding domain
 [KW] Irregular
 [KW] 3D
 [KW] LOW_COMPLEXITY 2.22 %
 [KW] COILED_COIL 8.54 %

SEQ MDTMMLNVRNLFQVLRVVEILSEGNEVQFIQLAKDFEDFRKKWQRTDHELKGYKDLLMK
 SEG
 COILSCCCCCCCCCCCC
 lrgp-
 SEQ AETERSALDKVLKHARNQVDVEIKRRQRAEADCEKLERQIQLIREMLMCDTSGSIQLSEE
 SEG
 COILS CC
 lrgp-
 SEQ QKSALAFLNRGQPSSSNAGNKRLLSTIDESGSILSDISFDKTDSESLDWDSSSLVKTFKLKKR
 SEG
 COILS

```

lrgp- .....
SEQ      EKRRTSRQFVDGPPGPVKKTRSIGSAVDQGNESIVAKTTVTVPNDGGPIEAVSTIETVP
SEG      .....
COILS    .....
lrgp- .....

SEQ      YWTRSRRTGTLPWNSTLNSRQLEPRTETDSVGT PQSNGGMRLLHDFVSKTVIKPESC
SEG      .....
COILS    .....
lrgp- .....

SEQ      VPCGKRIKFGKLSLKCRCRVVSHPECDRCPLPCIPTLIGTPVKIGEGMLADFVSQTSP
SEG      .....
COILS    .....
lrgp- .....

SEQ      MIPSIVVHCVNEIEQRGLTETGLYRISGCDRTVKELKEKFLRVKTVPLLSKVDDIHAICS
SEG      .....
COILS    .....
lrgp-    .CCHHHHHHHHHHHHHHTTTTTTTTCCCHHHHHHHHHHHHCCCCCG-GGCCCHHHHHH

SEQ      LLKDFLRNLKEPLLTFRNLRAFMEAAEITDEDNSIAAMYQAVGELPQANRDTLAFMLIHL
SEG      .....
COILS    .....
lrgp-    HHHHHHHHTTTTTTGGGHHHHHTTTT-CGGGHHHHHHHHHHHCCCHHHHHHHHHHHHHH

SEQ      QRVAQSPHTKMDVANLAKVFGPTIVAHAVPNPDVMTLQDIKRPKVVERLLSLPLEYWS
SEG      .....
COILS    .....
lrgp-    HHHHHHHHCCCHHHHHHHHGGGCC.....

SEQ      QFMMVEQENIDPLHVIENSNAFSTPQTPDIKVSLLGPVTTPEHQLLKT PSSSSLSQRVRS
SEG      .....
COILS    .....
lrgp-    .....

SEQ      TLTKNTPRFGSKSKSATNLGRQGNFFASPMKL
SEG      xxx.....
COILS    .....
lrgp-    .....

```

Prosites for DKFZphtes3_1c1.3

PS00001	212->216	ASN_GLYCOSYLATION	PDOC00001
PS00004	141->145	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	182->186	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	246->250	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	63->66	PKC_PHOSPHO_SITE	PDOC00005
PS00005	174->177	PKC_PHOSPHO_SITE	PDOC00005
PS00005	186->189	PKC_PHOSPHO_SITE	PDOC00005
PS00005	245->248	PKC_PHOSPHO_SITE	PDOC00005
PS00005	313->316	PKC_PHOSPHO_SITE	PDOC00005
PS00005	392->395	PKC_PHOSPHO_SITE	PDOC00005
PS00005	435->438	PKC_PHOSPHO_SITE	PDOC00005
PS00005	595->598	PKC_PHOSPHO_SITE	PDOC00005
PS00005	606->609	PKC_PHOSPHO_SITE	PDOC00005
PS00006	47->51	CK2_PHOSPHO_SITE	PDOC00006
PS00006	66->70	CK2_PHOSPHO_SITE	PDOC00006
PS00006	144->148	CK2_PHOSPHO_SITE	PDOC00006
PS00006	206->210	CK2_PHOSPHO_SITE	PDOC00006
PS00006	234->238	CK2_PHOSPHO_SITE	PDOC00006
PS00006	270->274	CK2_PHOSPHO_SITE	PDOC00006
PS00006	323->327	CK2_PHOSPHO_SITE	PDOC00006
PS00006	387->391	CK2_PHOSPHO_SITE	PDOC00006
PS00006	392->396	CK2_PHOSPHO_SITE	PDOC00006
PS00006	410->414	CK2_PHOSPHO_SITE	PDOC00006
PS00006	449->453	CK2_PHOSPHO_SITE	PDOC00006
PS00006	489->493	CK2_PHOSPHO_SITE	PDOC00006
PS00006	579->583	CK2_PHOSPHO_SITE	PDOC00006
PS00007	46->55	TYR_PHOSPHO_SITE	PDOC00007
PS00007	376->385	TYR_PHOSPHO_SITE	PDOC00007
PS00008	131->137	MYRISTYL	PDOC00008
PS00008	150->156	MYRISTYL	PDOC00008
PS00008	276->282	MYRISTYL	PDOC00008
PS00008	377->383	MYRISTYL	PDOC00008
PS00008	388->394	MYRISTYL	PDOC00008
PS00008	623->629	MYRISTYL	PDOC00008
PS00009	303->307	AMIDATION	PDOC00009

PS00479 287->336 DAG_PE_BINDING_DOMAIN PDOC00379

Pfam for DKFZphtes3_1c1.3

HMM_NAME Phorbol esters / diacylglycerol binding domain

HMM *HrFmrHTFrqPTWCDHCgeFIWGWgKQGYQCQnCgMCHKRCHelVPmm

H+F+ +T + P +C CG +I +GK ++C +C+++ H +C+ + P
Query 287 HDEFVSKTVIKPESCVPCKRI-KFGKLSLKCRDCRVVSHPECRDRCPLP 334

HMM C*

C
Query 335 C 335

DKFZphtes3_lgl3

group: intracellular transport and trafficking

DKFZp DKFZphtes3_lgl3 encodes a novel 1007 amino acid protein with similarity to human 256 kD golgin.

The new protein contains 7 leucine zippers and seems to be involved in protein-protein-interaction in the golgi apparatus. The very similar rat cpl51 shows haploid-specific transcription in mus musculus testis.

The new protein can find application in modulating protein traffic in the golgi apparatus, especially in human haploid germ cells.

similarity to 256 kD golgi, strong similarity to rat "cpl51"

21 exons encoded on AC004682

EST from a testis library, two mouse ESTs of a testis cDNA library, rat cpl51 shows haploid-specific transcription!
testis or haploid-specific transcription

Sequenced by DKFZ

Locus: map="16q22.2"

Insert length: 3405 bp

Poly A stretch at pos. 3394, polyadenylation signal at pos. 3373

```

1 GGGATAGGGG ATGTGGTTTG TTACAAAGGA TGAGTATTTT GATAGCTTCT
51 CATTCTTGA ACTATTCTGC AGGTTTATAA CAAAGCTCAG AAAATACTAA
101 AGGTTAAAGG AGAATTGAGA GCTGCCAAGG AAATGAAAGA TGAGGCGGGG
151 GAGAGAGACA GAGAAGTGAG CAGCCTGAAC AGCAAGCTGT TAAGCCTGCA
201 ACTTGACATC AAGAATCTGC ACGATGTCTG CAAGAGACAG AGGAAGACCT
251 TGCAGGACAA TCAGCTCTGC ATGGAGGAGG CAATGAACAG CAGCCACGAC
301 AAGAAGCAAG CACAGGCATT AGCATTGAGG GAGTCAGAGG TGGAATTTGG
351 GTCCAGTAAA CAGTGTCTATC TGAGACAACCT CCAGCAACTG AAGAAAAAAT
401 TGCTGGTCTT TCAACAAGAA CTGGAGTTTC ACACAGAGGA GTTGACAGCT
451 TCTTACTATT CTCTCCGCCA GTATCAGTCC ATCCTAGAGA AGCAGACTTC
501 CGACCTGGGT CTTCTGCACC ATCACTGCAA ACTGAAAGAA GATGAGGTGA
551 TTCTCTATGA GGAGGAAATG GGAAATCACA ACGAGAACAC AGGGGAGAAG
601 CTCATTTTGG CGCAGGAGCA ACTCGCCTTG GCCGGGGACA AGATCGCCTC
651 TCTAGAGAGG AGCTTAAACC TCTACAGGGA TAAATACCAG TCTTCCCTGA
701 GCAACATCGA GTTACTAGAA TGCCAAGTGA AGATGTTGCA GGGGGAACCT
751 GCGGGGATCA TGGGTCAGGA GCCTGAGAAC AAGGGTGATC ATTCAAAGGT
801 ACGGATATAC ACTTCTCCTT GCATGATTCA AGAGCATCAG GAGACTCAGA
851 AAGCAGCTGT TGAAGTCTGG CAAAAGGTCT CTCAACAGGA TGATCTCATT
901 CAAGAAGTTC GAAATAAGCT GGCCTGCAGT AACGCTTTGG TTCTGGAGCG
951 TGAAAAAGGCT TTGATAAAAC TACAAGCCGA TTTTGCTTCC TGTACAGCCA
1001 CCCACAGATA CCCTCCTAGC TCCTCAGAAG AGTGTGAAGA CATCAAAAAG
1051 ATRACTAAGC ACTTGCAAGG GCAGAAAGAC AGCCAGTGCC TGCAATGTGGA
1101 GGAGTACCAG AACCTGGTGA AGGATCTGCG CGTGGAACCTA GAGGCCGTGT
1151 CGGAACAGAA GAGAAACATC ATGAAGGACA TGATGAAGCT GGAGCTGGAG
1201 CTGCACGGAC TGGCGGAGGA GACATCTGCC CACATTGAGA GGAAGGATAA
1251 GGACATCACC ATCCTGCAGT GCCGGCTGCA GGAGCTGCAG CTGGAGTTCA
1301 CCGAGACCCA AAAGCTCACT TTGAAGAAAG ACAAGTTCCT CCAAGAGAAA
1351 GATGAGATGC TGCAAGAGCT GGAGAAGAAA CTGACACAGG TTCAGAACAG
1401 CCTCCTGAAA AAGGAGAAGG AGCTGGAGAA GCAGCAGTGC ATGGCCACAG
1451 AACTTGAAAT GACAGTCAAG GAGGCTAAGC AGGACAAGTC CAAGGAGGCG
1501 GAGTGCAAGG CCCTGCAGGC TGAGGTCCAG AAGCTGAAGA ACAGTCTCGA
1551 AGAGGCCAAG CAGCAGGAGA GGCTGGCTGC TCAGCAAGCA GCCCAGTGCA
1601 AAGAAGAGGC TGCACCTGCC GGCTGTCACC TGGAGGACAC CCAGAGGAAA
1651 CTGCAGAAGG GTCTCCTCCT GGACAAGCAG AAGGCAGACA CCATCCAGGA
1701 ACTACAGAGA GAACTTCAGA TGCTGCAGAA GGAGTCTCTG ATGGCTGAGA
1751 AGGAACAAAC CTCCAACAGA AAACGGGTGG AGGAGCTGTC ATTAGAACTC
1801 TCTGAAGCCC TGAGGAAGCT TGAAAATTCA GACAAGGAAA AGAGGCAGCT
1851 TCAGAAAGAC GTGGCTGAGC AGGATATGAA AATGAATGAC ATGCTTGATC
1901 GTATCAAGCA CCAGCACAGG GAGCAAGGCT CCATCAAATG CAAGTTAGAA
1951 GAAGATCTTC AGGAGGCCAC AAAGCTTCTG GAGGACAAAC GGGAGCAGTT
2001 GAAGAAGAGC AAAGAGCATG AGAAGCTGAT GGAGGGAGAA CTTGAAGCTT
2051 TCGCGGAGGA ATTTAAAAAG AAAGACAAGA CGTTGAAAGA GAATTCCAGA
2101 AAGTTGGAGG AAAAAATGA GAATCTCCGA GCAGAGCTAC AGTGTGTGTC
2151 TACACAAGTG GAATCCTCTC TCAACAAATA CAACACCAGC CAGCAAGTCA
2201 TCCAAGACTT GAATAAAGAG ATAGCCCTTC AGAAGGAGTC CTTAATGAGC
2251 CTGCAGGCCC AGCTGGACAA AGCTCTGCAG AAGGAGAAGC ACTATCTCCA
2301 GACTACCATC ACCAAAGAAG CCTATGATGC ATTATCCCGG AAGTCAGCCG
2351 CTGCCAGGGA TGACCTGACA CAAGCCCTCG AGAAGCTCAA TCACGTGACC
2401 TCAGAGACAA AGAGCCTGCA GCAAAGCTTG ACACAGACCC AAGAGAAGAA

```

```

2451 AGCTCAGCTG GAAGAGGAAA TCATTGCTTA TGAGGAAAGG ATGAAAAAGC
2501 TCAATACCGA ATTAAGAAAA CTGCGGGGCT TCCACCAGGA GAGTGAGCTG
2551 GAGGTGCACG CCTTTGACAA GAAGCTAGAG GAGATGAGCT GCCAGGTGCT
2601 GCAGTGGCAG AAGCAACACC AGAATGACCT CAAGATGCTG GCAGCCAAAG
2651 AGGAGCAGCT CAGGGAGTTC CAGGAGGAGA TGGCCGCCTT AAAAGAGAAC
2701 CTCCTTGAGG ACGATAAGGA GCCCTGTGTC CTGCCCCAGT GGTCTGTGCC
2751 CAAAGACACC TGTAGGCTCT ACCGAGGAA TGATCAGATT ATGACCAACT
2801 TGGAGCAATG GGC AAAACAG CAGAAGGTCG CCAATGAGAA ACTAGGAAAC
2851 CAGCTCCGAG AGCAGGTGAA CTACATTGCC AAGCTGAGTG GCGAAAAGGA
2901 CCACCTCCAC AGTGTAATGG TCCACTTGCA GCAGGAAAC AAGAAGCTGA
2951 AGAAGGAGAT AGAAGAGAAG AAGATGAAAG CCGAGAACAC AAGGCTATGC
3001 ACCAAAGCCC TAGGCCCGAG CAGAACGGAG TCCACACAGA GGGAGAAAGT
3051 GTGCGGCACC TTGGGCTGGA AGGGGTTGCC CCAGGATATG GGTCAAAGAA
3101 TGGACCTCAC CAAGTACATC GGGATGCCCC ACTGCCCGGG TTCCTCATAC
3151 TGCTAGAATC CACATCTAGC CCTGAGCAGC ATTTCCACGG GTGTTTCTTC
3201 AGAGGACAGT GAGTTCCAG CCCTCCCTCT CTCTTGACCT GGATCAGCTC
3251 TTACAGGAGT ATATCACGGT CCCAGCCTAT TTTGCAAGAC ACTAACTTTT
3301 GTTGAGTTTT GTCCACTTCC TGCCATGGAG TGAGCTTTAG AACCATACTA
3351 CCATCTCCAG GCCCAAACCTC TGAATAAAG ACATGAGCAT GAGCAAAAAA
3401 AAAAA

```

BLAST Results

Entry AC004682 from database EMBLNEW:

Homo sapiens Chromosome 16 BAC clone CIT987SK-A-259H10, complete sequence.

Score = 1291, P = 0.0e+00, identities = 265/272

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 133 bp to 3153 bp; peptide length: 1007

Category: similarity to known protein

Prosites motifs: LEUCINE_ZIPPER (83-105)

LEUCINE_ZIPPER (90-112)

LEUCINE_ZIPPER (97-119)

LEUCINE_ZIPPER (104-126)

LEUCINE_ZIPPER (403-425)

LEUCINE_ZIPPER (410-432)

LEUCINE_ZIPPER (918-940)

```

1 MKDEAGERDR EVSSLNSKLL SLQLDIKNLH DVCKRQRKTL QDNQLCMEEA
51 MNSSHDKKQA QALAFEESEV EFGSSKQCHL RQLQLKKKL LVLQQELEFH
101 TEELQTSYYS LRQYSILEK QTSDLVLLHH HCKLKEDEVI LYEEEMGNHN
151 ENTGEKLHLA QEQLALAGDK IASLERSLNL YRDKYQSSLS NIELLECOVK
201 MLQGELGGIM QGEPENKGDH SKVRIYTPC MIQEHQETQK RLSEVWQKVS
251 QDDLIQELR NKLACSNALV LEREKALIKL QADFASCTAT HRYPPSSSEE
301 CEDIKKILKH LQEQKDSQCL HVEEYQNLVK DLRVELEAVS EQKRNMKDM
351 MKLELDLHGL REETSAHIER KDKDITILQC RLQELQLEFT ETQKLTLLKD
401 KFLQEKDEML QELEKKLTQV QNSLLKKEKE LEKQQCMATE LEMTVKEAKQ
451 DKSKEAECKA LQAEVQKLKN SLEEAQQER LAAQQAQCK EEAALAGCHL
501 EDTORKLQKG LLLDKQKADT IQELQRELQM LQKESMAEK EQTSNRKRVE
551 ELSLELSEAL RKLNSDKEK RQLQKTVAEQ DMKMNDMLDR IKHQHREQGS
601 IKCKLEEDLQ EATKLLDKR EQLKKSKEHE KLMEGELEAL RQEFKKKDKT
651 LKENSRLLEE ENENLRAELQ CCSTQLESSL NKYNTSQQVI QDLNKEIALQ
701 KESLMSLQAG LDKALQKEKH YLQTTITKEA YDALSRSKSA CQDDLTQALE
751 KLNHVTSETK SLOQSLTQTO EKKAQLEEEI IAYEERMKKL NTELRKLGRF
801 HQESELEVHA FDKKLEEMSC QVLQWQKHQ NDLKMLAAKE EQLREFQEEM
851 AALKENLLED DKEPCCLPQW SVPKDTCLRY RGNDQIMTNL EQWAKQKVA
901 NEKLGNQREL QVNYIAKLSG EKDHLSVMV HLQQENKKLK KEIEEKKMKA
951 ENTRLCTKAL GPSRTESTQR EKVCGTLGWK GLPQDMGQRM DLTXYIGMPH
1001 CPGSSYC

```

BLASTP hits

Entry HS417401_1 from database TREMBL:

product: "trans-Golgi p230"; Human trans-Golgi p230 mRNA, complete

cds.

Score = 411, P = 3.9e-34, identities = 212/862, positives = 420/862

Entry SCINTANA_1 from database TREMBL:

Saccharomyces cerevisiae integrin analogue gene, complete cds.

Score = 404, P = 6.2e-34, identities = 199/897, positives = 423/897

Entry HS6802_2 from database TREMBL:

gene: "MYH9"; product: "dJ6802.2"; Homo sapiens DNA sequence from PAC 6802 on chromosome 22. Contains apolipoprotein L, myosin heavy chain, ESTs, CA repeat, STS and GSS.

Score = 404, P = 1.9e-33, identities = 231/1028, positives = 469/1028

Entry AF092090_1 from database TREMBL:

product: "cp151"; Rattus norvegicus cp151 mRNA, partial cds.

Score = 2523, P = 3.0e-262, identities = 506/733, positives = 611/733

Alert BLASTP hits for DKFZphtes3_lg13, frame 1

TREMBL:HSGOLGIN_1 product: "256 kD golgin"; H.sapiens mRNA for golgin,
N = 1, Score = 411, P = 4.4e-34

TREMBL:HS417401_1 product: "trans-Golgi p230"; Human trans-Golgi p230
mRNA, complete cds., N = 1, Score = 411, P = 4.5e-34

TREMBL:SCINTANA_1 Saccharomyces cerevisiae integrin analogue gene,
complete cds., N = 1, Score = 404, P = 7.1e-34

>TREMBL:HSGOLGIN_1 product: "256 kD golgin"; H.sapiens mRNA for golgin
Length = 2,185

HSPs:

Score = 411 (61.7 bits), Expect = 4.4e-34, P = 4.4e-34

Identities = 212/816 (25%), Positives = 420/816 (51%)

```
Query: 145 EMGNHNEN-TGEKLHLAQEQALAGDKIASLERSLNLYRDKYQSSLSNIELLECQVKMLQ 203
      +M + E+ G L +EQL ++ +ERSL+ YR KY ++ ++L+ + K LQ
Sbjct: 119 DMDSEAEDLVGNSDSLNEQLI---QRLRRMERSLSSYRGKYSSELVTAYQMLQREKKKLQ 175

Query: 204 GELGGIMGQEPENKGDHSHKVRITYTSPCMIQEHQETQKRLSEVWQ-KVSQQDDLIQELRNK 262
      G I+ Q D S RI +Q Q+ +K L E + + ++D I L+ +
Sbjct: 176 G----ILSQSQ----DKSLRRIAEELREELQMDQAKKHLQEEFDASLEEKDQYISVLQTQ 227

Query: 263 LAC-----SNALVLEREKALIKLQADFASCTATHRYPPSSSEEC-ED--IKKILKHLQE 313
      ++ + + ++ K L +L+ A P S E ED K L+ LQ+
Sbjct: 228 VSLLKQRLRNGPMNVDLKPLPQLEPQ-AEVFTKEENPESDGEPPVEDGTSVKLTLETLOQ 286

Query: 314 QKDSQ-----CLH-VEEYQNLVKDLRVELEAVSEQKRNIMKDMMKLELDLHGLREETSA 366
      + Q C ++ ++ L E EA+ EQ ++++ K++ DLH + E+T
Sbjct: 287 RVKQENLLKRCKETIQSHKEQCTLLTSEKEALQEQLDERLQELEKIK-DLH-MAEKTKL 344

Query: 367 HIERKDKDITILQCRQLQELQLEFTEQKLTLLKDKFLQEKDEMLQELEKKLTQV--QNSL 424
      + +D I Q Q+ + ET++ + + L+ K+E + +L ++ Q+ Q
Sbjct: 345 ITQLRDAKNLIEQLE-QDKGMVIAETKR---QMHEMLEMKEEIEIAQLRSRIKQMTTQGE 400

Query: 425 LKKEKELEKQOCMALETEMTVKEAKQDKSKEAECKALQAEVQKLNKSLLEAKQQERLAAQ 484
      L+++KE + ++ ELE + A+ K++EA K L+AE+ + ++E+ ++ER++ Q
Sbjct: 401 LREQKE-KSERAAFELEKALSTAQ--KTEBARRK-LKAEMDEQIKTIEKTSEEERISLQ 456

Query: 485 QA-AQCKEEAA-LAGCHLEDTQRKLQKGLLLDKQKADTIQELQRELQMLQKESMAEKEQ 542
      Q ++ K+E + E+ KLQK L +K+ A QEL ++LQ ++E E+ +
Sbjct: 457 QELSRVKQEVVDVMKKSSEEQIAKLQK--LHEKELARKEQELTKKLQTRERE--FQEQMK 512

Query: 543 TSNRRKVEELSLELSEALRKLKLENSDKEKRQLQKT--VAEQDMKMNDMLDRIKHQHREQGS 600
      + K E L++S+ + E+ E+ +LQK + E + K+ D+ +
Sbjct: 513 VALEKSQSEY-LKISQEKEQQESLALLEELELQKKAILTESENKLRDLQQAETRYRTRILE 571

Query: 601 IKCKLEEDLQEATKILLED-----KREQLKKSKEHEKLMEG---ELEALR-QEFKKKDKTL 651
      ++ LE+ LQE +D + E+ K +KE ++E ELE+L+ Q+ + L
Sbjct: 572 LESSLEKSLQENKNQSKDLAVHLEAKNKHKNKEITVMVEKHKTELESKHKQDALWTEKL 631

Query: 652 KENSRLKEEENENLRAELQCCSTQLESSL-NKYNTSQQVIQDLNKE----IALQKESLMS 706
      + ++ + E E LR + C + E+ L +K Q I+++N++ + +++ L S
Sbjct: 632 QVLKQYQTEMEKLRK---CEQEKETLLDKKEIIFQAHIEEMNEKTELEKLDVKQTELES 688

Query: 707 LQAQLDKALQKEKHLYLT--TITKEAYDALSRKSAACQDDLTQALEKLNHVSETKSLQ 764
      L ++L + L K +H L+ ++ K+ D + ++ A D+ Q V S K +
```

Sbjct: 689 LSSELSEVL-KARHKLEELSVLKDQTDKMKQELEAKMDE--QKNHHQQQVDSIIKEHEV 745

Query: 765 SLTQTQEKKAQLEEEIIAYEERMKKLNTLRLRGFQHESELEVHAFDCKLEEMSCQVLQ 824
S+ +T+ KA L+++I E +K+ + L++ + + E ++ + +L++ S ++

Sbjct: 746 SIQRTE--KA-LKDQINQLELLLKERDKHLKEHQAHVENLEADIKRSEGELOQASAKLDV 802

Query: 825 WQKQHNDLKLAAKEEQLEFQEEMAALKENLLEDDKEPCCLPQW-----SVPKDT-C-R 878
+Q +Q+ A EQ + ++E++A L++ LL+ + E L + + + KD C

Sbjct: 803 FQS-YQS-----ATHEQTKAYEEQLAQLQKLLDLETERILLTKQVAEVEAQKKDVCTE 855

Query: 879 LYRGNDQIMTNLEQWAKQKQVANERKLNQLEQVNYIAKLS-GEKDLHLSVMVHLQQENK 937
L Q+ ++Q KQ +K+ + QV Y +KL G K+ + + ++EN

Sbjct: 856 LDAHKIQVQDLMQQLEKQNSEMEQKVKSLT--QV-YESKLEDGNKEQEQTQILVEKENM 912

Query: 938 KLK-KEIEEKKMAENTRLCTK 958
L+ +E ++K+++ +L K

Sbjct: 913 ILQMRGQKKEIEILTQKLSAK 934

Score = 338 (50.7 bits), Expect = 3.1e-26, P = 3.1e-26
Identities = 216/953 (22%), Positives = 468/953 (49%)

Query: 2 KDEAGERDRE--VSSLNS-KLL-SLQLDIKNLHDVCKRQRKTLQDN-QLCM-----EEAM 51
K+E E D E V S K L +LQ +K ++ KR ++T+Q + + C +EA+

Sbjct: 260 KEENPESDGEPPVEDGTSVKTLETQQRVKQENLLKRCKETIQSHKEQCTLLTSEKEAL 319

Query: 52 NSSHDKKQAQALAFESESEVEFGSSKQCHLRQ----LQQLK--KKLLVLQOELFHTTEELQ 105
D++ + ++ + + LR ++QL+ K +++ + + H E L+

Sbjct: 320 QEQLDERLQELEKIKDLHMAEKTKLITQLRDAKNLIEQLEQDKGMVIAETKRQMH-ETLE 378

Query: 106 TSYYSRLQYQSILEKQTSIDLVLHHCKLKEDEVILYEEEMGNHNENTGEKLHLAQEQL- 164
+ Q +S +++ T+ L K K + E E +T +K A+ +L

Sbjct: 379 MKEEEIAQLRSRIQMTTQGEELREQ-KEKSERAFAFELEKAL---STAQKTEEARRKLK 434

Query: 165 ALAGDKIASLERSLNYRDKYQSSLSNI--ELLEQVQKMLQGLGGIMGQEPENKGDHSH 222
A ++I ++E++ R Q LS + E+++ K + ++ + Q+ K K

Sbjct: 435 AEMDEQIKTIEKTSEEERISLQQLSRVKQEVVDVMKKSSEEQIAKL--QKLHEKELARK 492

Query: 223 VRIYTSPCMIQEHQETQKRLSEVWQKVSQDDLIQELRNKLACSNAVLREKALIKLQA 282
+ T +E +E Q+++ +K SQ + L ++ + +L LE ++LQ

Sbjct: 493 EELTKKLQTRE-REFQEQMKVALEK-SQSEYL--KISQEQEQESLALAE---LELQK 544

Query: 283 DFASCTATHRYPPSSSEECEDIKKILKHLQEOKDSQCLHVEEYQNLVKDLRVELEAV-SE 341
A T + +E E + + L+ + ++E +N KDL V LEA ++

Sbjct: 545 K-AILTESENKLRLDQEAETRYRILELESSLEKS---LQENKNQSKDLAVHLEAEKKN 600

Query: 342 QKRNIKMMKLELDLHGLREETSABIERKDKDITI-LQCRLQELQLEFTETQKLTLLKD 400
+ I + K + +L L+ + A K + + Q +++L+ E E +K TL KD

Sbjct: 601 HNKEITVMVEKHKTELESLSKHQDQDALWTEKLQVLKQYQTEMEKLR-EKCEQEKETLLKD 659

Query: 401 K-----FLQEKDEM-LQELEKLTQVQNSLLKKEKELEKQCCMATELEMTVKEAQDKS 453
K ++E +E L++L+ K T+++ SL + E+ K + E E++V + + DK

Sbjct: 660 KEIIFQAHIEEMNEKTLEKLDVKTQTELE-SLSSELSEVLKARHKEE-ELSVLKDQTDKM 717

Query: 454 K-EAECKALQAEVQKLNLSLEAKQERLAAQQAQC-KEEAALAGCHLEDQTKLQKGL 511
K E E K + + ++ ++ ++ Q+ + K++ L++ + L++

Sbjct: 718 KQELEAK-MDEQKNHHQQQVDSIIKEHEVSIQRTEKALKDQINQLELLLKERDKHLKEHQ 776

Query: 512 L-LDKQKADTIQELQRELQMLQKESMAEKEQTSNRKRVEELSLELSEALRKLENSDKEK 570
++ +AD I+ + ELQ + + + Q++ ++ + +L++ +KL + + E+

Sbjct: 777 AHVENLEAD-IKRSEGELOQASAKLDVFSYQSATHEQTKAYEEQLAQLQKLLDLETER 835

Query: 571 RQLQKTVAEQDMKMDM--LD--RIKHQHREQGSIK--CKLEEDLQEATKLEDKREQL 623
L K VAE + + D+ LD +I+ Q Q K ++E+ ++ T++ E K E

Sbjct: 836 ILLTKQVAEVAQKKDVCTELDAHKIQVQDLMQQLEKQNSEMEQKVKSLTQVYESKLEDG 895

Query: 624 KKSKEHEK--LMEGELEALRQEFKKDKTLKENSRLKEENENLRAELQCCSTQLESSLN 681
K +E K L+E E L+ +K K ++ ++KL + +++ + T+ ++

Sbjct: 896 NKEQEQTQILVEKENMILQMRGQK-KEIEILTQKLSAKEDSIHILNEEYETKFNQEK 954

Query: 682 KYNTSQVQIDLNKEIALQKESLMSLQAQLDKALQKEKHYLQTTITKEAYDALSRSKSAAC 741
K +Q +++ + + K+ L+ +A+L K L E L+ + ++ ++A + A

Sbjct: 955 KMEKVKQKAKEMQETL---KKKLLDQEAQLKEL--ENTALELSQKEKQFNAKMLEMAQA 1009

Query: 742 QD-DLTQALEKLNHVTSSETKSLOQSLTQTQEKKAQLEEEIIAYEERMKKLNTLRLRGF 800
++ A+ +L T++ + ++ SLT+ + +L + I +E KKLN + +L+

Sbjct: 1010 NSAGISDAVSRLE--TNQKEQIE-SLTEVHRR--ELNDVISIWE---KKLNQQAELQEI 1061

Query: 801 HQESELEVHAFDCKLEEMSCQVLQW--QKQHNDLKLAAKEEQLEFQEEMAALKENLL 858
H E+++ ++++ E+ ++L + +K+ N ++ KEE +++ + L+E L

Sbjct: 1062 H---EIQLQEKEQVAELKQKILLFGCEKEEMNK-EITWLKEEGVKQ-DTTLNELQEQLK 1116

Query: 859 EDDKEPCCLPQWSVPKDTCLRYRGNDQIMTNLEQ--WAKQQKVANEKLGNLREQVNYI- 915
 + L Q K L + + +L++ + ++Q V + L + + +V+ +
 Sbjct: 1117 QKSAHVNSLAQ-DETKLKAHLEKLEVDLNKSLKENTFLQEQVLKMLAEEDKRVSELT 1175

Query: 916 AKLSGEKDHLHSMVHLQQENKKLK-KEIEEKKMAE 951
 +KL + S+ ++ NK L+ K +E KK+ E
 Sbjct: 1176 SKLKTTDEEFQSLKSSHEKSNKSLKEDKSLEFKKLSSE 1212

Score = 337 (50.6 bits), Expect = 4.0e-26, P = 4.0e-26
 Identities = 215/951 (22%), Positives = 433/951 (45%)

Query: 10 REVSSLNSKLLSLQLDIKNLHDVCKRQRKTLQDNQLCMEAMSSHDKKQAALAFEESE 69
 +E + +++L L+ ++ K Q K L + EA + H+K+ + E+ +
 Sbjct: 560 QEAEYTRTRILELESSLEKSLQENKNQSKDLAVHL----EAEKNKHNEKIT--VMVEKHK 613

Query: 70 VEFSSKQCHLRQLQQLKKLLVLQQLLEFHTTEELQTSYSLRQYQSILEKQTSIDLVLH 129
 E S K H +Q +KL VL+Q+ + E+L+ Q + L K +++
 Sbjct: 614 TELESK--H-QQDALWTEKLQVLKQYQTEMEKLEK--CEQEKETLLKD-KEIIFQA 666

Query: 130 HHCKLKE---DEVILYEEEMGNHNENTGEKL---HLAQEQALAGDKIASLERSLNLRYD 183
 H ++ E +++ + + E+ + + E L H +E+L++ D+ +++ L D
 Sbjct: 667 HIEEMNEKTLEKLDVKQTELESLSSESEVLKARHKEEELSVLKDQTDKMKQLEAKMD 726

Query: 184 K----YQSSLNIEELLECCQVKMLQGE--LGGIMGOEPENKGDHKSRIYTSPCMIQEHQE 237
 + +K + +I + E +V + + E L + Q + K + ++ +
 Sbjct: 727 EQNHHQQQVDSI-IKEHEVSIQRTEKALKDQINQLELLKERDK-HLKEHQAHVENLEA 784

Query: 238 TQKRLSEVWQKVSQDDLIQELRNKLACSNALVLEREKALIKLQADFASCTATHRYPPSS 297
 KR Q+ S + D+ Q ++ ++ E+ L +LQ T R
 Sbjct: 785 DIRSEGEQQAASAKLDVFSYQS---ATHEQTKAYEEQLAQLQKLLDLE--TERIL--- 837

Query: 298 SECEDIKKILKHLQEOKDSQCLHVEEYQNLVLDLRELEAVSEQKRNIMKDMML-ELD 356
 + K + ++ QK C ++ ++ V+DL +LE + + +K + ++ E
 Sbjct: 838 -----LTKQVAEVEAQKKDVCTELDAHKIQVQDLMQQLEKQNSEMEQKVKSLTQVYESK 891

Query: 357 LH-GLREETSAHIERKDKDITILQCRL-QELQLEFTETQKLTLLKKDKF--LQEKDEM-LQ 411
 L G +E+ +K+ ILQ R Q+ ++E TQKL+ K+D L E+ E +
 Sbjct: 892 LEDGNKEQEQTQKILVEKENMILQMREGQKKEIEIL-TQKLSAKEDSIHILNEEYETKFK 950

Query: 412 ELEKKLTVQVNSLLK-----KEKELEKQCMATELEMTVKEAKQDKSKEAECKALQAEVQ 466
 EKK+ +V+ + K+K L+++ + ELE T E Q K K+ K L+ Q
 Sbjct: 951 NQEKMKKVKQAKEMQETLKKLLDQEAALKKELENTALELSQ-KEQFNAKMLEM-AQ 1008

Query: 467 KLKNSLEAKQERLAAQQAACKEEAALAGCHLEDTRKQLKGLLLDKQKADTIQELQR 526
 + +A RL Q Q + + L D +K L Q+A+ +QE+
 Sbjct: 1009 ANSAGISDAVS--RLETNQKEQIESLTVHRRELNDVISIWEKKL---NQQAEELEQEIH- 1062

Query: 527 ELQMLQKESMAEKEQT-----SNRKR--EELSLESEALRKLKLENSDKERQLQ 574
 E+Q+ +KE +AE +Q K + +E ++ L +L+ K+K
 Sbjct: 1063 EIQLQEKEQEVAELKQKILLFGCEKEEMNEKITWLKEEGVKQDITLNLQEQQLKQSAHV 1122

Query: 575 KTVAEQDMKMDMLDRIKHQHREQGSIKCKLEEDLQEQATKLEDKREQLKKSKEHEKLME 634
 ++A+ + K+ L++++ + L+E L E L E+ + ++ + K +
 Sbjct: 1123 NSLAQDETKLKAHLEKLEVDLNKSLKENTFLQEQVLKMLAEEDKRVSELTSLKTTD 1182

Query: 635 GELEALRQEFKKDKTLKENSRLKEEENENLRAELQCCSTQLESSLNKYNTSQQVIQDLN 694
 E ++L+ +K +K+L++ S + ++ +E L +L C + E+ L T++ + +
 Sbjct: 1183 EEFQSLKSSHEKSNKSLKEDKSLEFKKLSLELAQLDICCKKTEALLEA-KTNELINISS 1241

Query: 695 KEIALQKESLMSLQALDKALQKEHYLQTTITKEAYDALSRKSAACQDDLT----QALE 750
 K A+ + Q + K KE ++T E +A R+ Q+ L QA
 Sbjct: 1242 KTNAILSR-ISHQHRTTKV--KEALLIKTCTVSEL-EAQLRQLTEEQNTLNISFQQATH 1297

Query: 751 KLNHVTSSETSLQQLTQTEKKAQLEEEIAYEERMKKLN---TELK--LRGFHQESE 805
 +L ++ KS++ + +K L++E ++ + T+L+K + +
 Sbjct: 1298 QLEEKENQIKSMKADIESLVTEKEALQKEGGNQQAASEKESCITQLKKELSENINAVTL 1357

Query: 806 LEVHAFDKKLE--EMSCQVLQWQKHQNDLKLAAKEEQLEFQEMAALKENLLEDDKE 863
 ++ +KK+E +S Q+ Q QN + L+ KE + +++ K LL D +
 Sbjct: 1358 MKEELKKEKVEISSLSKQLTDLNLVQLQNSIS-LSEKAAISSLRKYDEEKCCELL-DQVQ 1415

Query: 864 PCCLPQWSVPKDTCLRYRGNDQIMTNLEQWAKQQKVANEKLGNLRE---QVNYIAKLSG 920
 ++ K+ D +W K+ + + N ++E Q+ +K +
 Sbjct: 1416 DLSFKVDTLSKEKISALEQVDDWSNKFSEWKKKAQSRFTQHNTVKELQIQLELKSKEAY 1475

Query: 921 EKDH-LHSMVHLQQENKK--LKKEIEEKKMAE 951
 EKD ++ + L Q+NK+ LK E+E+ K K E
 Sbjct: 1476 EKDEQINLLKEELDQQNKRFDCCLKGEMEDDKSKME 1510

Score = 332 (49.8 bits), Expect = 1.4e-25, P = 1.4e-25
 Identities = 209/953 (21%), Positives = 438/953 (45%)

Query: 1 MKDEAGERDREVSSLNSKLLSLQLDIKNLHDVCKRQRKTLQDNQLCMEEAMNS----SHD 56
 MK + E+ ++ L+ K L+ + + + + R+R+ + ++ +E++ + S +
 Sbjct: 470 MKKSSEEQIAKLQKLHEKELARK-EQELTKKLQTREREFQEQMKVALEKSQSEYKISQE 528

Query: 57 KKQAQALAFESEVEFGSSKQCHLRQLQQLKKLLVLQOELEFHTTEELQTSYSLRQYQS 116
 K+Q ++LA EE E++ K+ L + + KL LQOE E + + SL +
 Sbjct: 529 KEQESLALAELELQ----KKAILTESEN---KLRDLQQAETRYRTRILESSLEKSLQ 581

Query: 117 ILEKQTSIDLVLHHHCKLKEDE--VILYEE----EMGNHNENT--GEKLHLAQEQLALA 167
 + Q+ DL + K K ++ ++ E+ E H ++ EKL + ++Q
 Sbjct: 582 ENKNQSKDLAVHLEAKKNHNKEITVMVEKHKTELESKHKQDQDALWTEKLQVLKQYQTE 641

Query: 168 GDKIASL--ERSLNLYRDK---YQSSLS--NIELLECQVKMLQGELEGGIMGQEPENKGDH 220
 +K+ + L +DK +Q+ + N + LE ++ + Q EL + + E
 Sbjct: 642 MEKLREKCEQEKETLLKDKKIIIFQAHIEEMNEKTLE-KLDVKQTELESLSSELSEVLKAR 700

Query: 221 SKVRIYTPCMIQEHQETQKRLSEVWQKVSQDDLIQELRNKLACSNALVLEREKALIKL 280
 K+ S ++++ +T K E+ K+ +Q + Q+ + + + ++R+ +K
 Sbjct: 701 HKLEEELS--VLKD--QTKMKQLEAKMDEQKNHHQQVDSIIKEHEVSIQRTKALKD 756

Query: 281 QADFASCTATHR--YPPSSSEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEA 338
 Q + R + E+++ +K + + ++ +Q+ + +A
 Sbjct: 757 QINQLELLLKERDKHLKEHQAHVENLEADIKRSEGELQASAKLDVFSYQSATHEQTKA 816

Query: 339 VSEQKRNIMKDMMLLELDLHGLREETSALIERKDKDITILQCRQLQELQEFTEQKLTLLK 398
 EQ + + ++ LE + L ++ A +E + KD+ C EL + Q L +
 Sbjct: 817 YEEQLAQLQQLLDLETERILLTKQV-AEVEAQKQDV---CT--ELDAHQIQVQDLMMQ 869

Query: 399 KDKFLQEKDEMLQLEKKLTQVQNSLLKK-EKELEKQCCMATELEMTVKEAKQDKSKEAE 457
 +K + EM Q++ K LTQV S L+ KE E+ + + E E + + ++ + KE E
 Sbjct: 870 LEK---QNSEMEQKV-KSLTQVYESKLEDGKKEQEQTQKILVEKENMILQMRGQKKEIE 925

Query: 458 C--KALQAEVQKLKNSLEEAKQERLAAQQAQCKEAAAGCHLEDTRQK--LQKGLLL 513
 + L A+ + EE + + + ++ + K+A ++T +K L + L
 Sbjct: 926 ILTQKLSAKEDSIHILNEEYETKFKNQEKMEKVKQKAK---EMQETLKKLLDQEAAL 981

Query: 514 DKQKADTIQEL-QRELQMLQKESMAEKEQTSNRKRVEELSLELSEALRKLNSDKERQ 572
 K+ +T EL Q+E Q K MA+ V L E + L ++ +R+
 Sbjct: 982 KKELENTALELSQKEKQFNAKMLEMAQANSAGISDAVSRLETNQEKEQIESL--TEVHRRE 1039

Query: 573 LQKTVAEQDMKMNMDLDRIKHQHREQGSICKLEEDLQEQATKLEEDKREQLKKS----KE 628
 L ++ + K+N + ++ H Q K + +L++ L ++E++ K KE
 Sbjct: 1040 LNDVISIWEKKLNQQAELQEIHEIQLOKEQEVAELKQKILLFGCEKEEMNKEITWLKE 1099

Query: 629 HEKLMEGELEALRQEFKKKDKTLKENSRLKEEENENLRAELQCCSTQLESSLNKYNTSQ 688
 + L L+++ K+K + NS L ++ L+A L+ L SL + Q+
 Sbjct: 1100 EGVKQDITLNLQEQQLKQSAHV--NS--LAQDETCLKAHLEKLEVDLNSKSLKENTFLQE 1155

Query: 689 VIQDLNKEIALQKESLMSLOAQL---DKALQ--KEKHYLQTTITKEA---YDALSRSAA 740
 + +L K + L ++L D+ Q K H ++ + LS + A
 Sbjct: 1156 QLVELKMLAEEDKRKVSLETSKLTDEEFQSLKSSHEKSNKSLKEDKSLFKKLSSE-LA 1214

Query: 741 CQDDL-----TQAL-----EKLNHVTSETKSLQSLTQTEKKAQLEEEIIAYEERMKKL 790
 Q D+ T+AL E +N +S+T ++ ++ Q + ++E ++ +L
 Sbjct: 1215 IQLDICCKKTEALAEAKTNELINISSSKTNAILSRISHCQHRRTTKVKEALLIKTCTVSEL 1274

Query: 791 NTELRLKRGFQHESELEVHAFDKKLEEMSCQVLQWQKHQNDLKMLAAKEEQLEFQEE 850
 +LR+L + +LEE Q+ K + D++ L ++E L Q+E
 Sbjct: 1275 EAQLRQLTQEQNTLNISFQATHQLEEKENQI----KSMKADIESLVTEKEAL---QKEG 1327

Query: 851 AALKENLLEDDKEPCCLPQWSVPKDTCLRYRGNDQIMTNLEQWAKQKQVANEKLGQNLRE 910
 + +KE C + Q + K+ N +T +++ K++KV L QL +
 Sbjct: 1328 G--NQQAASEKESC-ITQ--LKKELSE----NINAVTLMKEELKEKKVEISSLSKQLTD 1378

Query: 911 ---QVNYIAKLSGEKDLHLSVMVHLQENKLLKKEIEEKKMAE 951
 Q+ LS ++ + S+ +E +L ++++ K +
 Sbjct: 1379 LNVQLQNSISLSEKAAISSLRKQYDEEKCELLDQVQDLSFKVD 1422

Score = 329 (49.4 bits), Expect = 2.9e-25, P = 2.9e-25
 Identities = 226/941 (24%), Positives = 444/941 (47%)

Query: 61 QALAFESEVE--FGSSKQCHLRQLQQLKKLLVLQOELEFHTTEELQTSYSLRQYQSIL 118
 Q L E+ +++ S+ LR++ +L+++L + QQ + EE S QY S+L
 Sbjct: 165 QMLQREKKKLQGILSQSQDKSLRRIAELREELQMDQQAQKHLQEEFDASLEEKDQYISVL 224

Query: 119 EKQTSIDLVLHHHCKLKEDEV-----ILYEEEMGNHNENT---GEKL---HLAQEQLALA 167
 + Q S L + + D + + + E+ EN GE + + L
 Sbjct: 225 QTQVSLKQLRLNGFPMNVDLKPLPQLEPQAEVFTKEENPESDGPVVEDGTSVKTLET 284

Query: 168 GDKIASLERSLNLYRDKYQSSLSNIELLECQVKMLQGELEGGIMGQEPENKGDHDKVRIYT 227

Sbjct: 285 ++ E L ++ QS LL ++ LQ +L + QE E D ++ QQRVKRQENLLKRCKETIQSHKEQCTLLTSEKEALQEQLDERL-QELEKIKD---LHMAE 340

Query: 228 SPCMIQEHQETQKRLSEVWQKVSQQDDLIQELRNKLACSNAVLEREKALIKLQADFASC 287
+I + ++ + ++ Q +I E + ++ L ++ E + +L++

Sbjct: 341 KTKLITQLRDAKNLIEQLEQDKGM---VIAETKRQM--HETLEMKEEE-IAQLRSRIKQM 394

Query: 288 TATH---RYPSSSEEC--EDIKKILKHLQEOKDSQCLHVEEYQNLVKDL-----RVE 335
T R SE E+++K L Q+ +++ E +K + R+

Sbjct: 395 TTQGEELREQKEKSERAAFELEKALSTAQKTEEARRKLKAEMDEQIKTIEKTSEEERIS 454

Query: 336 LEA-VSEQKRNMKDMMKL--ELDLHGLREETSASIERKDKDITILQCRLQELQLEFTET 392
L+ +S K+ ++ D+MK E + L++ + RK++++T +LQ + EF E

Sbjct: 455 LQQLSRVKQEVV-DVMKKSSEEQIAKLQKLHEKELARKEQLTK---KLQTREREFQEQ 510

Query: 393 OKLTLKKDKFLQEKDEMLQLEKKLTQVQNSLLKKEKELEKQCCMATELEMTVKEAKQDK 452
K+ L+K + E ++ QE E+ Q SL +E EL+K+ + TE E +++ +Q+

Sbjct: 511 MKVALEKSQ--SEYKISQEKEQ-----QESLAELELELQKKAIL-TESENKLRLDQEQ- 561

Query: 453 SKEAECKALQAEVQKLKNSLEEAKQER-----LAAQQAQCKEEAALAGCHLEDTQR-K 506
++ +L+ E L+ SL+E K Q + L A++ KE + H + + K

Sbjct: 562 AETYRTRILELE-SSLEKSLQENKNQSKDLAVHLEAEKNKNKEITVMVEKHKTELESK 620

Query: 507 LQKGLLLDKQKADTIQELQRELQMLQKESMAEKEQTSNRKRVEELSLELSEALRK-LEN 565
Q+ L ++ Q+ Q E++ L +E EKE K + + E K LE

Sbjct: 621 HQQDALWTEKLQVLKQYQTEMEKL-REKCEQEKETLLKDKKEII-FOAHIEEMNEKTLEK 678

Query: 566 SDKEKRQLQKTVAEQDMKMNMDLDRIKHQHREQSGI-KCKLEEDLQEA-TKLEDKR--E 621
D ++ +L+ +E ++++L + +H+ E+ S+ K + ++ QE K+ E K +

Sbjct: 679 LDVKQTELESLSSE----LSEVL-KARHKLEELSVLKQDTDKMKQELEAKMDEQKNHHQ 733

Query: 622 QLKKS--KEHEKMEGELEALRQEFKKDKTLKENSRLKEEEN---ENLRAELQCCSTQL 676
Q S KEHE ++ +AL+ + + + LKE + L+E ENL A+++ +L

Sbjct: 734 QQVDSIIKEHEVSIQRTEKALKDQINQLELLKDKHLKEHQAHVENLEADIKRSEGEL 793

Query: 677 ESSLNKYNTSQQVIQDLNKEIALQKESLSLQAQLDKALQKEKHYLQTTITKEAYDALSR 736
+ + K + Q +++ +E L LQ +L L+ E+ L TK+ + ++

Sbjct: 794 QQASAKLDVFSYQSATHEQTKAYEEQLAQLQKKL-LDLETERILL----TKQVAEVEAQ 848

Query: 737 KSAACQD-----DLTQALEKLNHVTSSETKSLQOQSLTQTEKKAQ--LEEEIIAYEE 785
DL Q LEK N SE + +SLTQ E K + +E+ +

Sbjct: 849 KKDVTCTELDAHKIQVQDLMOQLEKQN---SEMEQVKSLTQVYESKLEDGNKEQEQTQKI 905

Query: 786 RMKKLNTELRLRGFHOESELEVHAFDKKLEEMSCQVL--QWQKHQNDLKMLAAKEEQ 843
++K N L+ G Q+ E+E+ +E S +L +++ + +N K + +++

Sbjct: 906 LVEKENMILQMREG--QKKEIEILTQKLSAKEDSIHILNEEYETKFKNQEKMKMKVQKA 963

Query: 844 REFQEEAALKENLEDDKEPCLPQWSVPKDTCLRYRGNDQIMTNLEQWAKQKQV---- 899
+E QE LK+ LL+ + + L + + L + Q + + A+

Sbjct: 964 KEMQE---TLKKLLDQEAQ---LKK-ELENTALELSQKEQFNAKMLEMAQANSAGISD 1016

Query: 900 ANEKLGNLQREQVNYIAKLSG-EKDHLHSMVH-LQQENKKLKK--EIEEKKMKAENTRL 955
A +L +EQ+ + ++ E + + S+ L Q+ ++L++ EI+ ++ + E L

Sbjct: 1017 AVSRLETNQKEQIESLTFVHRRLENDVISIWEKKLNQQAEELEIHEIQLQEKEQEVAE 1076

Query: 956 CTKALGPSRTESTQREKVCGLGWKGLPQD 985
K L E + K L +G+ QD

Sbjct: 1077 KQKIL-LFGCEKEEMNKETWLKEEGVKQD 1105

Score = 326 (48.9 bits), Expect = 6.0e-25, P = 6.0e-25
Identities = 220/907 (24%), Positives = 444/907 (48%)

Query: 67 ESEVEFGSSKQCHLRQLQQLKKLLVLQLEFHTTEELQTSYYSLRQYQSILE---KQTS 123
E+E G+S + QL Q +++ EL T+Y L++ + L+ Q+

Sbjct: 123 EAEDLVGNSDSLKEQLIQLRRMERSLSSYRGKSELVTAYQMLQREKKKLQGILSQSQ 182

Query: 124 DLVLLHHCKLKEDEVILYEEEMGNHNTGEKHLAQEQLALAGDKIASLERSLNLYRD 183
D L +L+E+ + +++ H + E+ + E+ I+ L+ ++L +

Sbjct: 183 DKSL-RRIAELREE--LQMDQQAKKHLQ---EEFDASLEE---KDQYISVLQTVSLLKQ 233

Query: 184 KYQSSLSNIELLECQVKMLQGELGGIMQE-PENKG-----DHSKVR-IYTSPCMIQEHQ 236
+ ++ N+++L+ + L+ + +E PE+ G D + V+ + T ++ +

Sbjct: 234 RLRNGPMNVDLK-PLPQLEPQAEVFTKEENPESDGEPPVEDGTSVKTLETQQRVKRQE 292

Query: 237 ETQKRLSEVWQKVSQQDDLIQELRNKLACSNAVLEREKALIKLQADFASCTATHRYPPS 296
KR E Q +Q L+ K A L ER + L K++ D T

Sbjct: 293 NLLKRCKETIQSHKEQCTLLTS--EKEALQEQLD-ERLQELEKIK-DLHMAEKTCLIT-- 346

Query: 297 SSEECEDIKKILKHLQEOKDSQCLHVEEYQNLVKDLRVELEAVSEQKRNMKDMMKLELD 356
+ D K +++ L++ K + E + + + L ++ E ++ Q R+ +K M +

Sbjct: 347 ---QLRDAKNLIEQLEQDKGM--VIAETKRQMHEHTEMKEEEIA-QLRSRIKQMTTQGE 400

Query: 357 LHGLREETS-AHIERKDKDITILQCRLE---LQLEFTETQKLTLLKDKFLQEKDEMLQ 411
 L +E++ A E +K ++ Q + +E L+ E E K T++K +E+ + Q
 Sbjct: 401 LREQKEKSERAAFELEKALSTAQ-KTEEARRLKAEMDEQIK-TIEKTSE-EERISLQQ 457

Query: 412 ELEKLTQVQNSLLKK-EKELEKQCCMATELEMTVKEAKQDKSKEAECKALQAEVQKLN 470
 EL + +V + + K E+++ K Q + E E+ KE Q+ +K+ + + + Q +K
 Sbjct: 458 ELSRVKQEVVDVMKKSSEEQIAKLQKLH-EKELARKE--QELTKKLQTREREFQEQ-MKV 513

Query: 471 SLEEAKQERLAAQQAQCKEEAALAGCHLEDTQRKLQ-KGLLLD-KQKADTIQELQREL 528
 +LE++ Q E L Q + +E AL L+ + + L D +Q+A+T + EL
 Sbjct: 514 ALEKS-QSEYLIKISQEKQESLAELELELQKKAILTESENKLRDLQQAETRYRTRILEL 572

Query: 529 QMLQKESMAEKEQTSNRKRVEELSLELSEALRKLENS-DKEKRQLQKTVAEQDMKMNDM 587
 + E S+ E + S V L E ++ +++ +K K +L+ +QD +
 Sbjct: 573 ES-SLEKSLQENKNQSKDLAVH-LEAEKNKHKEITVMVEKHKTELESKHHQQDALWTEK 630

Query: 588 LDRIKHQHR-EQSSIKCKLEEDLQEA TKLLEDKRE--QLKKSKEHEKLMEGELEALRQEF 644
 L +K Q++ E ++ K E QE LL+DK Q + +EK +E +L+ + E
 Sbjct: 631 LQVLKQQYQTEMEKLREKCE---QEKETLLKDKKEIFQAHEEMNEKTLE-KLDVKQTEL 686

Query: 645 KKKDKTLKE--NSR-KLEENENLRAELQCCSTQLESSLNKY-NTSQVVIQDLNKE--IA 698
 + L E +R KLEE L+ + +LE+ +++ N QQ + + KE ++
 Sbjct: 687 ESLSSELSEVLKARHKLEELSVLKDOTDKMKQELEAKMDEQKNHHQQQVDSIIKEHEVS 746

Query: 699 LQK-ESLSMLQA-QLDKAL-QKEKHYLQTTITKEAYDALSRK-----AACQDDLTQAL 749
 +Q+ E + Q QL+ L +++KH + E +A ++S A+ + D+ Q+
 Sbjct: 747 IQRTEKALKDQINQLELLKRDKHLKEHQAHVENLEADIKRSEGELOQASAKLDVFSY 806

Query: 750 EKLNVHTSETKSLQQLSTQTEKKAQLEEEIIAYEERMKKLNTLRLKRGFHESELEVH 809
 + H +TK+ ++ L Q Q+K LE E I +++ ++ + + + +++V
 Sbjct: 807 QSATH--EQTKAYEEQLAQLQQLLDLETERILLTKQVAEVEAQKDKDVCTELDAHKIQVQ 864

Query: 810 AFDKKLEEMSCQVLQWQKQHON--DLKMLAAKEEQLREFQEEMAALKENLL----EDDK 863
 ++LE+ + ++ Q K + K+ +EQ E +++ KEN++ E K+
 Sbjct: 865 DLMQQLKQNSEMEQKVKSLSLQVYESKLEDGNKEQ--EQTKQILVEKENMILQMREGQKK 922

Query: 864 PC-CLPQ-WSVPKDTCLRYRGNDQIMTNLE-QWAKQKQVANE--KLGNLQREQV-NYIAK 917
 L Q S +D+ + N++ T + Q K +KV + ++ L++++ + AK
 Sbjct: 923 EIEILTQKLSAKEDSIHIL--NEEYETKFKNQEKMKVKQKAKEMQETLKKKLLDQEA 980

Query: 918 LSGEKDLHLSVMVHLQQENKKLKEIEEKKMKAENTRLCTKALGPSRTESTQREKV 973
 L K L + + L Q+ K+ ++ E M N+ + A+ SR E+ Q+E++
 Sbjct: 981 L---KKELENTALELSQKEKQFNAKMLE--MAQANSAGISDAV--SRLETNQKEQI 1029

Score = 318 (47.7 bits), Expect = 4.4e-24, P = 4.4e-24
 Identities = 184/827 (22%), Positives = 405/827 (48%)

Query: 1 MKDEAGERDREVSSLSKLLSLQLDIKNLHDVCKRQRKTLQDNQLCMEAMNSSHDKK-Q 59
 ++ E G + + S S + L+ ++ + ++ L++ ++ + D Q
 Sbjct: 1323 LQKEGGNQQAASEKESCITQLKKELSENINAVTLMKEELKEKKVEISSLSKQLTDLNVQ 1382

Query: 60 AQ-ALAFEESEVEFGSSKQCHLRQLQQLKKLLVLQQELFHELTQTSYYS-LRQYQS- 116
 Q +++ E E S + +Q + K +LL Q+L F + L S L Q
 Sbjct: 1383 LQNSISSEKAAISSLR---KQYDEEKCELLDQVQDLSFKVDTLSKEKISALEQVDDW 1438

Query: 117 ---ILE-KQTSDLVLLHHHCKLKEDEVILYEEEMGNHNENTGEKLHLAQELALAGDKIA 172
 E K+ + H +KE ++ L + + ++ E+++L +E+L +
 Sbjct: 1439 SNKFSEWKKKAQSRFTQHONTVKELQIQLELKSKEAYEKD--EQINLLKEELDQQNKRFD 1496

Query: 173 SLERSLNLRYDKYQSSLSNIEL-LECQVKMLQGELEGGIMGQEP-ENKGDHSHKVRIYTPC 230
 L+ + + K + SN+E L+ Q + EL + Q+ E + + ++ Y
 Sbjct: 1497 CLKGEMEDDKSKMEKKESNLETTELKSTARIM-ELEDHITQKTIEISLNEVLKNYNQQK 1555

Query: 231 MIQEHQETQKRLSEVQKVSQDDLIQELRNKLACSNAVLEREKALIKLQADFASCTAT 290
 I EH+E ++L + ++D+ ++E K+ L LE + +K + +
 Sbjct: 1556 DI-EHKELVQKLQHFQELGEEKDNRVKEAEKI-----LTLENQVYSMAELETKKKELE 1609

Query: 291 HRYPPSSSEECEDIKKILKHLQEQKDSQCLHVE-EYQNLVKDLRVELEAVSEQKRNMKD 349
 H S+E E+K + L+ + ++ ++ + + + ++ +L + E+K ++
 Sbjct: 1610 HVNLSVKSKE-EELKALEDRLESESAAKLAELKRKAQKIAAIKKQLLSQMEEK----EE 1664

Query: 350 MMKLELDLHGLREETS-AHIERKDKDITILQCRLEQLQLEFTETQKL--TLKDKFLQEKD 407
 K + H E + + + +++++ IL+ +L+ ++ +ET + + K E++
 Sbjct: 1665 QYKKGTEH--LSELNTKLQEREREVHILEEKLKSVSSQSETLIVPRSAKNVAAYTEQE 1722

Query: 408 EM-----LQEL-EKLTQVQNSLLKKEKEL-----EKQCCMATELEMTVK-EAKQDKSKE 455
 E +Q+ E+K++ +Q +L +KEK L EK++ +++ EM + + + K +
 Sbjct: 1723 EADSGQGVQKTYEEKISVLQRNLTEKEKLLQRVGQEKEETVSSHFMRCQYQERLIKLEH 1782

Query: 456 AECKAL--QAEVQKLKNSLEEAKQERLAAQQAQCK--EEAALAGCHLEDTQRKLQKGL 511

AE K Q + L + LEE ++ L Q + + + A +LE+ +QK L
 Sbjct: 1783 AEAKQHEDQSMIGHLQEELEEKNKYSLIVAQHVKEGGKNNIQAQNLNVFDDVQKTL 1842
 Query: 512 LLDKQKADTIQELQRELQMLQKESSMAEKEQTSNRKRVEELS--LELSEALRKLENSDKE 569
 ++K T Q L+++++ L +S + +++ +R +EEL+ E +AL++++ +K
 Sbjct: 1843 ---QEKELTQCILEQKIKEL--DSCLVRQKEV-HRVEMEELTSKYEKLQALQMDGRNKP 1896
 Query: 570 KRQLQKTVAEQD---MKMNDMLDRIKHQHREQSGIKCKLEEDLQEATKLEEDKREQLKK- 625
 L++ E+ + +L ++ QH + E + Q+ K + ++ L+
 Sbjct: 1897 TELLEENTEEKSKSHLVQPKLLSNMEAQHNDFKLAGAEREKQKLGEIVRLQKDLRML 1956
 Query: 626 SKEHEKMEGELEALRQEFKKDKTLKENSRLKEEENENLRAELQCCSTQLESSLNKYNT 685
 KEH++ ELE L++E+ + E K+++E E+L EL+ ST L+ + ++NT
 Sbjct: 1957 RKEHQ---ELEILKKEYDQ-----EREKIKQEEDL--ELKHNT-LKQLMREFNT 2003
 Query: 686 S-QQVIQDLNKEIALQKESLSLQAQLDKALQKEKHYLQTTITKEAYDALSRKSAACQDD 744
 Q Q+L I ++A+L ++ Q+E + L I E D L R +A ++
 Sbjct: 2004 QLQKEQELEMTIKETINKAQEVEAELESHQEETNQLLKKIA-EKDDDLKR-TAKRYEE 2061
 Query: 745 LTQALEKLNHVTSKSLQOSLTQTQEKKAQ-LEEEIAYEERMK--KLNTLRKLRGFH 801
 + A E+ +T++ + LQ L + Q+K Q LE+E + + +L T+L +
 Sbjct: 2062 ILDAEE--EMTAKVRDLQTQLEELQKKYQKLEQEENPGNDNVTIMELQTQLAQKTTLI 2119
 Query: 802 QESELEVHAFDKKLEEMSCQVLQWQK 827
 +S+L+ F +++ + ++ ++K
 Sbjct: 2120 SDSKLKEQEFREQIHNLDRKKYK 2145
 Score = 316 (47.4 bits), Expect = 7.1e-24, P = 7.1e-24
 Identities = 213/977 (21%), Positives = 454/977 (46%)
 Query: 4 EAGERD-REVSSLSKLLSLQLD-IKNLHDVCKRQKTLQDNQLCMEEAMNSSHDKKQAQ 61
 E R+ +V S+ K L+ Q + ++ +H++ + Q K + +L + + ++ +
 Sbjct: 1034 EVHRRRLNDVISIWEKLNQQAEEELQEIHAI-QLQKEQEVAELKQKILLFGCEKEEMNK 1092
 Query: 62 ALAFEESEVEFGSSKQCHLRQLQ-QLKKLL---VLQOE--LEFHTTEELQTSYYSLRQY 114
 + + + E G + L +LQ QLK+K + Q E L+ H E+L+ +
 Sbjct: 1093 EITWLKEE---GVKQDTTLNELQEQQLKQSAHVNSLAQDETFLKAHLEKLEVDLNSKSLKE 1149
 Query: 115 QSILEKQTSOLVLLHHHCKLKEDEV---ILYEEEMGNHNENTGEKHLHAEQLALAGDKI 171
 + L++Q +L +L K K E+ + +E +++ EK + + E +L K+
 Sbjct: 1150 NTFLEQLVLELKLAEEDKRVSELTSLKTTDEEFQSLKSSHEKSNSKLEDSLEFKKL 1209
 Query: 172 AS-LERSINLYRDYQSSLS--NIELLECQVKMLQGEELGIMGQEPENKGDHKSVMRIYTS 228
 + L L++ K ++ L EL+ L I +++ K +
 Sbjct: 1210 SEELAIQLDICCKTEALAEAKTNELINISSKTNAISRI--SHCQHRRTKVKKEALLIK 1267
 Query: 229 PCMIQEHQ-----ETQKRLSEVQKVSQQ-DDLIQELRNKLACSNAVLEREKALIKL 280
 C + E + E Q L+ +Q+ + Q ++ +++ A +LV E+E L
 Sbjct: 1268 TCTVSELAQLRQLTEEQNTLNISFQQATHQLEEKENQIKSMKADIESLVTEKEA---L 1323
 Query: 281 QADFASCTATHRYPPSSSEECEDIKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAVS 340
 Q + + + S E C I ++ K L E ++ L EE +K+ +VE+ ++S
 Sbjct: 1324 QKEGGN---QQQAASEKESC--ITQLKKELSENINAVTLMKEE----LKEKKVEISSLS 1373
 Query: 341 EQKRNIMKMMKLELDLHGLREETSABIERKDKDITILQCRQLQEL--QLEFTEQKLT-L 397
 +Q ++ + + L S+ ++ D++ L ++Q+L +++ +K++ L
 Sbjct: 1374 KQLTDLNVQLQN-SISLSEKAAISSLRQYDEEKCELLDQVQDLSFKVDTLSKEKISAL 1432
 Query: 398 KK-DKFLQEKDEMLQLEKKLTQVQNSLLKKEKELEKQCCMATELEMTV---KEAKQDKS 453
 ++ D + + E ++ + + TQ QN++ + + +LE + A E + + KE ++
 Sbjct: 1433 EQVDWNSKFSWKKAQSRFTQHONTVKELQIQLELKSKEAYEKDEQINLLKEELDQON 1492
 Query: 454 KEAECKALQAEVQKLKNSLEEAKQERLAAQQAQCKEEAALAGCHLE-DTQRKLQKGLL 512
 K +C + E K K +E+ + L +Q A + E + +E ++ ++ K
 Sbjct: 1493 KRFDCLKGEMEDDKSKMEKKESNLETLSQTARIMELEDHITQKTIEIESLNEVLKNY- 1551
 Query: 513 LDKQKADTIQELQRELQMLQKESSMAEKEQTSNRKRVEELSLELSEALRKLENSDKEKRO 572
 ++QK +EL ++LQ Q+ + +++ L ++ +LE KE
 Sbjct: 1552 -NQQKDIEHKELVQKLQHFQELGEEKDNRVKEAEKILTLENQVYSMKAELETKKKELEH 1610
 Query: 573 LQKTVAEQDMKMNMDLRIKHQHREQ-GSIKCKLEEDLQEATKLL---EDKREQLKSK 627
 + +V ++ ++ + DR++ + + +K K E+ + K L E+K EQ KK
 Sbjct: 1611 VNLSVSKKEELKALEDRLSESAAKLAELKRKAQKIAIAIKQLLSQMEKEEQYKKG 1670
 Query: 628 EHEKMEGELEALRQEFKKDKTLKENSRLKEE-ENENL---RAELQCCSTQLESSLNK 682
 E EL QE +++ L+E + +E ++E L A+ T+ E + ++
 Sbjct: 1671 ESHL---SELNTKLQEREREVHILEEKLKSVESQSETLIVPRSAKNVAAYTEQEEADSQ 1727
 Query: 683 ---YNTSQQVIQDLNKEIALQKESLSLQAQLDKALQKEKHYLQTTITKEAYDALSRKSA 739
 T ++ I L + + +KE L+ Q +K H+ +E L A
 Sbjct: 1728 GCVQKTYEEKISVLQRNLT-EKEKLLQRVQ-EKEETVSSHFMRCQYQERLIKLEHAEA 1785

Query: 740 ACQDDLTQALEKLNHVTSET--KSLQQLTQTQEKKAQLEEEIIAYEERMKKLNTELRLK 797
 +D Q++ + H+ E K+ + SL Q + + + I ++ ++ + +++K
 Sbjct: 1786 KQHED--QSM--IGHLQEELEEKNNKYSILVAQHVEKEGGKNNIQAQKNLENVFDVQKT 1841

Query: 798 RGFHQESELEVHAFDCKLEEM-SCQVLQWQKQHNDLKMALAAKEEQLEFQEEMAALKEN 856
 QE EL ++K++E+ SC V Q ++ H+ +++ L +K E+L+ Q+ K
 Sbjct: 1842 L---QEKELTCQILEQKIKELDSCLVQRQ-KEVHRVEMEELTSKYELQALQMDGRNKPT 1897

Query: 857 -LLEDDKEPCCLPQWSVPKDTCLRYRGNDQIMTNLEQWAKQKQVANKEKLGNLREQVNYI 915
 LLE++ E PK + ++ + L A+++K +KLG ++ +
 Sbjct: 1898 ELLEENTEESKSHLVQPKLLSNMEAQHNDLEFKLAG-AEREK---QKLGKEIVRLQKDL 1953

Query: 916 AKLSGE-KDHLHSMVHLQQENK-KLKKEIEEKKMAENTRLCTKALGPSRTESTQREK 972
 L E + L + QE + K+K+E E+ ++K +T + + T+ Q+E+
 Sbjct: 1954 RMLRKEHQEILEIKKEYDQEREKIKQEEDLELKHNT--LQQLMREFNTQLAQKEQ 2010

Score = 301 (45.2 bits), Expect = 2.9e-22, P = 2.9e-22
 Identities = 221/952 (23%), Positives = 441/952 (46%)

Query: 1 MKDEAGERDREVSSLNSKLLSLQLDIKNLHDVCKRQRKTLDQNL---CMEEAMNSSHD- 56
 +K A E R+VS L SKL + ++L ++ K+L+D L + E + D
 Sbjct: 1160 LKMLAEEDKRKVELTSKLTDEEFQSLKSSHEKSNKSLEDKSLFKKLSEELAIQLDI 1219

Query: 57 --KKQAQALAFESE-VEFGSSK-QCHLRQLQQLKKLLVLQOELEFHT---EELQTSYY 109
 KK L + +E + SSK L ++ + + +++ L T EL+
 Sbjct: 1220 CCKKTEALLEAKTNELINISSKTNAILSRISHCQHRTTKVKALLIKTCTVSELEAQLR 1279

Query: 110 SLRQYQSILEKQTSDDLVLHHHCKLKEDEVILYEEEMGNHNENTGEKLHLAQE---QLAL 166
 L + Q+ L H + KE+++ ++ EK L +E Q
 Sbjct: 1280 QLTEEQNTLNISFQAT---HQLEKENQIKSMKADI---ESLVTEKEALQKEGNGQQA 1333

Query: 167 AGDKIASLERSLNLYRDKYQSSLSNIELEECQVKMLQGEELGGIMGQEPENKGDHDKSVRIY 226
 A +K E + + + +++ + L++ ++K + E+ + Q + V++
 Sbjct: 1334 ASEK----ESCITQLKKELSENINAVTLMKEELKEKKVEISSLSKQLTD-----LNVQLQ 1384

Query: 227 TSPCMIQEHQETQKRRLSEVWQKVSQQDDLIQELRNKLACSNAVLEREKALIKLQADFAS 286
 S + ++ + ++ + D +Q+L K+ + L E+ AL ++ D+++
 Sbjct: 1385 NSISLSEKAAISSLRKQYDEEKCELLDQVQDLSFKV---DTLSKEKISALEQVD-DWSN 1440

Query: 287 CTATHRYPPSS--SEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKD-----LRVE-LE 337
 + + S ++ +K++ L E K + +E NL+K+ R + L+
 Sbjct: 1441 KFSEWKKKASRFTQHONTVKELQIQL-ELKSKEAYEKDEQINLLKEELDQONKRFDCIK 1499

Query: 338 AVSEQQRNIM-KMMKLELDLHGLRE---ETSAHIERKDKDITILQCRLOEL-QLEFTET 392
 E ++ M K LE +L E HI +K +I L L+ Q + E
 Sbjct: 1500 GEMEDDKSRMEKKESNLETELKSQTARIMELEDHITQKTIEISLNEVLKNYNQKQDIEH 1559

Query: 393 QKLTLLKDKFLQ---EKDEMLQELEKLTQVQNSLLKKEKELEKQQCMATELEMTVKEAK 449
 ++L K F + EKD ++E E+K+ ++N + + ELE ++ + ++VK
 Sbjct: 1560 KELVQKLQHFQELGEEKDNVRKEAEKILTLENQVYSMAELETKKKELEHVNLSVK--- 1616

Query: 450 QDKSKEAECKALQAEVQKLKNSLEEAKQERLAAQQAQKKEAALAGCHLEDTQKRLQK 509
 SKE E KAL+ ++ S + + +R A Q+ A K++ +E+ + + +K
 Sbjct: 1617 ---SKEEELKALEDRLES--ESAAKLAELKRAEQKIAAIKKQLL---SQMEEKEEQYK 1668

Query: 510 GLLLDKQKADT-IQELQRELQMLQKESMAEKEQTSNRKRVEELSLELSEALRKLNSDK 568
 G + +T +QE +RE+ +L+++ E Q+ + S + A + E +D
 Sbjct: 1669 GTESHLSELNTKLQEREREVHILEEKLKSVESQSETL--IVPRSAKNVAAYTEQEEADS 1726

Query: 569 E----KRQLQK-TVAEQDMKMND-MLDRIKHQHQSGSIKCKLEEDLQEATKLEDKREQ 622
 + K +K +V +++ + +L R+ Q +E+ ++ E Q +L+ K E
 Sbjct: 1727 QGCVQKTYEEKISVLQRNLTEKEKLLQVRG-QEKEE-TVSSHFEMRCQYQERLI--KLEH 1782

Query: 623 LKKSKEHE-KLMGEL-EALRQEFKKKDKTLKENSRLKEEENLRAELQCCSTQLESSL 680
 + +K+HE + M G L E L ++ KK + ++ K E N++A+ LE
 Sbjct: 1783 AE-AKQHEDQSMIGHLQEELEEKNNKYSILVAQHVEK-EGGKNNIQAQ-----QNLE--- 1832

Query: 681 NKYNTSQQVIQDLNKEIALQKESLMSLQAQLDKAL--QKEKHYLQTTITKEAYDALSR-K 737
 N ++ Q+ +Q+ KE+ Q L +LD L QKE H ++ Y+ L +
 Sbjct: 1833 NVFDDVQKTLQE--KELTCQ--ILEQKIKELDSCLVQRQKEVHRVEMEELTSKYELQALQ 1888

Query: 738 SAACQDDLTQALEKLNHVTSETKSLQQLTQTQEKKAQ-LEEEIIAYEERMKKLNTEL-- 794
 ++ T+ LE+ S++ +Q L E+ LE ++ E +KL E+
 Sbjct: 1889 QMDGRNKPTELLEENTEESKSHLVQPKLLSNMEAQHNDLEFKLAGAEREKQKLGKEIVR 1948

Query: 795 --RKLGRGFHQESELEVHAFDCKLEEMSCQVLQWQKQHNDLKMALAAKEEQLEFQEEMAA 852
 + LR +E + E+ K+ ++ + ++ Q+Q +LK + ++ +REF ++A
 Sbjct: 1949 LQKDLRLRKEHQEILEIKKEYDQEREKIK-QEQEDLELKHNTSLKQLMREFNTQLAQ 2007

Query: 853 LKENLLEDDKEPCCLPQWSVPKDTCLRYRGNDQIMTNLEQWAKQKQVANKEKLGNLREQV 912

Sbjct: 2008 KEQELEMTIKETINKAQ-EVEAELESH---QEETN--QLLK--KIA-EKDDDLKRTAK 2057
 Query: 913 NYIAKLSGKDLHLSVMVHLQKENKKLKEIEKKMKAEN 952
 Y L ++ + + + LQ + ++L+K+ ++K + EN
 Sbjct: 2058 RYEEILDAREEEMTAKVRDLQTQLEELQKKYQQKLEQEEN 2097

Score = 300 (45.0 bits), Expect = 3.7e-22, P = 3.7e-22
 Identities = 195/961 (20%), Positives = 435/961 (45%)

Query: 1 MKDEAGERDREVSSLSKLLSLQLDIKN--LHDVCKRQRKTLQDNQLCMEEAMSSHDKK 58
 +KD+ + +N K L +LD+K L + + L+ +EE ++ D+
 Sbjct: 657 LKDKIEIFQAHIEEMNEKTLE-KLDVKQTELESLSSELSEVLKARHK-LEEELSVLKQDT 714
 Query: 59 QAQALAFEESEVEFGSSKQCHLRQLQQLKKLLV-LQEELEFHTTEELQTSYSLRQYQSI 117
 +E E + K H +Q+ + K+ V +Q+ + +++ L++
 Sbjct: 715 DKMK---QELEAKMDEQKNHHQQVDSIIKEHEVSIQRTKALKDQINQLELLKERDKH 771
 Query: 118 LEKQTSDLVLLHHCKLKEDEVILYEEEMG---NHNENTGEKLHLAQEQLALAGDKIASL 174
 L++ + + L K E E+ ++ ++ T E+ +EQLA K+ L
 Sbjct: 772 LKEHQAHVENLEADIKRSEGELOQASAKLDVFSYQSATHEQTKAYEEQLAQLQKLLDL 831
 Query: 175 ERSNLNRYDKYQSSLSNIELLECQVKMLQGGELGGIMGQ-EPENKGDHDKSVRIYTSPCMIQ 233
 E L + + + + + + + + + + +M Q E +N KV+ T
 Sbjct: 832 ETERILLTKQVAEVEAQKKDVCTELDAHKIQVQDLMQQLEKQNSEMEQKVSLTQ-VYES 890
 Query: 234 EHQETQKRLSEVWQKVSQDDLIQELRN---KLACSNALVLEREKALIKLQADFASCTA 289
 + ++ K + Q + ++++I ++R ++ + +E ++ L ++ +
 Sbjct: 891 KLEDGNEQEQTQKILVEKENMILQMREGQKKEIEILTQKLSAKEDSIHILNEEYET--- 947
 Query: 290 THRYPPSSSEECEDIKKLKLQEQKDSQCLHVEEYQNLVKDLRVELEAVSEQKRNIMKD 349
 ++ + ++ E +K+ K +QE + L E L K+L +S++++
 Sbjct: 948 --KFK-NQEKMEKVKQAKEMQETLKKLLDQEA--KLKKELENTALELSQKEKQFNAK 1002
 Query: 350 MMKL-ELDLHGLREETS-A-HIERKDKDITILQCRLOELQLEFTETQKLTLLKDKFLQEKD 407
 M+++ + + G+ + S +K++ + +EL + +K ++ + LQE
 Sbjct: 1003 MLEMAQANSAGISDAVSRLTNQKEQIESLTVHRRELNDVISIWEKKLNQAEELQEIH 1062
 Query: 408 EM-LQELEKKTQVQNSLLK---KEKELEKQQCMATE---LEMTVKEAKQD-KSKEAEC 458
 E+ LQE E+++ ++ +L +++E+ K+ E + T+ E ++ K K A
 Sbjct: 1063 EIQLQKEQEVAELKQKILLFGCEKEEMNKEITWLKEEGVKQDTTLNELQEQKQKSAHV 1122
 Query: 459 KALQAEVQKLKNSLEEAQQRERLAAQQAQCKEEAALAGCHLEDTRKRLQKGLLLDKQKA 518
 +L + KLK LE+ + + ++ +E+ E+ +RK+ + L K K
 Sbjct: 1123 NSLAQDETCLKAHLEKLEVDLNSLSKENTFLQEQVELKMLAEEDKRKVS--LTSKLT 1180
 Query: 519 DTIQELQRELQMLQKESMAEKEQTSNRKRVEELSLELSEALRKLENSDKERQLQKTVA 578
 T +E Q +K + E + +K EEL+++L +K E + K + +
 Sbjct: 1181 -TDEEFQSLKSSHEKSNSKSLKEDKSLFKLSEELAIQLDICKKTEALLEAKTN--ELIN 1237
 Query: 579 EQDMKMNDMLDRIKH-QHREQGSICKLEEDLQEAATKLEEDKREQLKKSKEHEKLMGEL 637
 K N +L RI H QHR K++E L T + + QL++ E + +
 Sbjct: 1238 ISSSKTNAISRISHCQHRTT-----KVKEALLIKTCTVSELAQLRQLTEEQNTLNISF 1292
 Query: 638 EALRQEFKKKD---KTLKENSRLKEEENENLR-----AELQCCSTQLESSL---- 680
 + + ++K+ K++K + L E E L+ +E + C TQL+ L
 Sbjct: 1293 QQATHQLEEKENQIKSMKADIESLVTEKEALQKEGGNQQAASEKESCITQLKKELSENI 1352
 Query: 681 NKYNTSQQVIQDLNKEIALQKESLSLQALDKALQ-KEKHYLQTTITKEAYDALSRKSA 739
 N ++ ++ + EI+ + L L QL ++ EK +++ K+ YD +
 Sbjct: 1353 NAVTLMKEELKEKKVEISSLSKQLTDLNVQLQNSISLSEKAAISSLRKQ-YDEEKCELL 1411
 Query: 740 ACQDDTLQALEKLN-HVTSETKSLQSSLTQTQEKKAQLEEEIIAYEERMKKLNTCLR-KL 797
 DL+ ++ L+ S + + + E K + + ++ +K+L +L K
 Sbjct: 1412 DQVQDLSFKVDTLSKEKISALEQVDDWSNKFSEWKKKAQSRFTQHNTVKELQIQLELKS 1471
 Query: 798 RGFHQSELEVHAFDKKLEEMSCQVLQWQKHQNDLKMLAAKEEQLR-EFQEEMAALKEN 856
 + +++ E +++ ++L++ + + + ++D + KE L E + + A + E
 Sbjct: 1472 KEAYEKDE-QINLLKEELDQONKRFDCLGEMEDDKSKMEKKESNLETELKSTARIME- 1529
 Query: 857 LLEDDKEPCCLPQWSVPKDTCLRYRGNDQIMTNLEQWAKQKQVANERLGNLREQVNYIA 916
 LED + + T + N+ ++ N Q QK K +L +++ +
 Sbjct: 1530 -LEDH-----ITQKTIEIESLNE-VLKNNYQ----QKDIEHK---ELVQKLQHFQ 1570
 Query: 917 KLSGEKDH----LHSMVMVHLQKENKKLKEIEKKMKAENTRLCTKA 959
 +L EKD+ ++ L+ + +K E+E KK + E+ L K+
 Sbjct: 1571 ELGEEKDNRVKEAEKILTLENQVYSMAEALETKKKELEHVNLSVKS 1617

Score = 298 (44.7 bits), Expect = 6.1e-22, P = 6.1e-22
 Identities = 207/886 (23%), Positives = 412/886 (46%)

Query: 47 MEEAMNSSHDKQAQALAFEESEVEFGSSKQCHLRQLQQLKKLLVLQQLQEFHTEELQT 106
 + E N + + Q EE E + S K ++ L + LQ+E +
 Sbjct: 1281 LTEEQNTLNISFQQATHQLEEKENQIKSMKA----DIESLVTEKEALQKEGGNQQAASE 1336

Query: 107 SYSLRQYQSILEKQTSIDLVLHHHCKLKEDEVILYEEEMGNHNENTGEKLHLAQEQLAL 166
 + Q + L + + + L+ K K+ E+ +++ + N + L++++ A
 Sbjct: 1337 KESCITQLKKELSENINAVTLMKEELKEKKVEISSLSKQLTDLNVQLQNSISLSEKEAA- 1395

Query: 167 AGDKIASLERSLNLYRDKYQSSLSNIELLECQVKMLQGLGGIMGQEPENKGDHDKSVRIY 226
 I+SL + Y ++ L ++ L +V L E + Q + S+ +
 Sbjct: 1396 ----ISSLRKQ---YDEEKCELLDQVQDLSFKVDTLSKEKISALEQVDDWSNKFSEWK-K 1447

Query: 227 TSPCMIQEHQETQKRLS-----EVWQKVSQDDLIQEL--RNK-LACSNALVLE--- 272
 + HQ T K L E ++K Q + L +EL +NK C + +
 Sbjct: 1448 KAQSRTFQHNTVRELQIQLELKSKEAYEKDEQINLLKEELDQONKRFDCCLKGEMEDDKS 1507

Query: 273 -REKALIKLQADFASCTAT----HRYPPSSSEECEDIKKILKHLQEQKDSQCLHVEEYQN 327
 EK L+ + S TA + + E E + ++LK+ +QKD E++
 Sbjct: 1508 KMEKKESNLETELKSQTARIMELEDHITQKTIEIESLNEVLKNYNQKQDI-----EHKE 1561

Query: 328 LVKDLRVELEAVSEQKRNMKDMKLELDLHGLREETSATIERKDKDI--TILQCRLOEL 385
 LV+ L+ + + E+K N +K+ + L L A +E K K++ L + +E
 Sbjct: 1562 LVQKLQ-HFQELGEEKDNRVKEAEEKILTLENQVYSMAELETKKKELEHVNLSVKSKEE 1620

Query: 386 QLEFTTETQKLTLLKKDKFLQEKDEMLQLEKLLTQVQNSLLKKEKELEKQCCMATELEMTV 445
 +L+ E + L+ + + E+ ++E+K+ ++ LL + +E E+Q TE ++
 Sbjct: 1621 ELKALEDOR---LESES-AAKLAELKRKAQKIAAIIKKQLLSQMEEKEEYKKGTESHLSE 1676

Query: 446 KEAQDKSKEAECKALQAEVQKLKNSLEEAKQQRERLAAQQAQCK-EAAALAGCHLEDTQ 504
 K + +E E L+ +++ ++S E R A AA + EEA GC + +
 Sbjct: 1677 LNTKLQE-REREVHILEEKLKSVESSESQSETLIVPSAKNVAAYTEQEEADSGGCVCQKTYE 1735

Query: 505 RKLQKGLLLDKQKADTIQELQRELQMLQKESMAEKEQTSNRKRVEELSLELSEALRKLE 564
 K+ +L + + + LQR Q +KE +++ + R + +E ++L A K
 Sbjct: 1736 EKIS---VLQRNLTEKEKLLQVRGQ--EKEETVSSHFE--RCQYQERLIKLEHAEAKQH 1788

Query: 565 NSDKERQLQKTVAEQDMKMNMDLDRIKHQHQREQ--SIKCK--LE---EDLQ-----E 611
 LQ+ + E++ K + ++ +H +E G +I+ K LE +D+Q E
 Sbjct: 1789 EDQSMIGHLQEELEKNKYSLIV--AQHVEKEGGKNNIQAQNLNVFDDVQKTLQKEKE 1846

Query: 612 AT-KLEDKREQLKKSKEHEKLMEG-ELEALRQEFKKDKTLKENS-----KLEENENL 665
 T ++LE K ++L +K + E+E L +++K + + R +L EEN
 Sbjct: 1847 LTCQILEQKIKELDSCLVRQKEVHRVEMEELTSKYELQALQQMDGRNKPTELLEENTEE 1906

Query: 666 RAELOCCSTQLESSLN-KYNTSQQVIQDLNKEIALQKESLMSLQAQLDKALQKEKHLYQT 724
 +++ ++L S++ ++N + + +E + ++ LQ L + L+KE H +
 Sbjct: 1907 KSKSHLVQPKLLSNMEAQHNDLEFKLAGAEREKQKLGKEIVRLQKDL-RMLRKE-HQOEL 1964

Query: 725 TITKEAYDALSRKSAACQDDLTQALEKLNHVTSETKSLQQSLTQTQEKKAQLEEEIAYE 784
 I K+ YD R+ Q+ + LE L H ++ + +++ TQ +K+ +LE I +
 Sbjct: 1965 EILKKEYDQ-EREKIKQE--EDLE-LKHNSLTKLMREFNTQLAQKEQLEMTI---K 2017

Query: 785 ERMKKLNTELRLRGFHFQSELEVEHAFDKKLEEMSCQVLQWQKHQNDLKMALAAKEEQLR 844
 E + K +L HQE E + KK+ E + + K+++ ++L A+EE++
 Sbjct: 2018 ETINKQEVEAEELSHQE---ETNQLKKIAEKDDDLKRTAKRYE---EILDAREEEMT 2071

Query: 845 EFQEEMAALKENLLEDDKEPCCLPQWSVP-KDTCRLYRGNDQIMTNLEQWAKQKQVANKE 903
 ++ EL + ++ L Q P D + ++ T L Q K +++ K
 Sbjct: 2072 AKVRDLQTLEELQKQYQK--LEQEENPGNDNVTIM---ELQTQLAQ--KTTLISDSK 2123

Query: 904 LGNQ-LREQVNYIA-KLSGEKDHLHSMV-HL 932
 L Q REQ++ + +L + +++++ V HL
 Sbjct: 2124 LKEQEFREQIHNLEDRLLKKEKNVYATTVGHL 2155

Score = 280 (42.0 bits), Expect = 5.2e-20, P = 5.2e-20
 Identities = 209/938 (22%), Positives = 432/938 (46%)

Query: 3 DEAGERDREVS-SLNSKLLSLQLDIKN-LHDVC-KRQRKTLQDNQLCMEEAM-NSSHDKK 58
 ++ ++ +E+ +L KLL + +K L + + +K Q N +E A NS+
 Sbjct: 957 EKVKQKAKEMQETLKKKLLDQEAALKKELENTALELSQKEKQFNKMLEMAQANSAGISD 1016

Query: 59 QAQALAFEESEVEFGSSKQCHLRQLQQLKKLLVLQQLQEFHTEELQTSYSLRQYQSIL 118
 L + E + S + H R+L + + + +++L EELQ + ++ +
 Sbjct: 1017 AVSRLETNQKE-QIESLTVHRRELNDV---ISIWKKLNQQAELQ-EIHEIQLQEK-- 1069

Query: 119 EKQTSIDL--LLHHHCKLKE-DEVILYEEEMGNHNENTGEKLHLAQEQLALAGDKIASLE 175
 E++ ++L +L C+ +E ++ I + +E G + T +L +Q + + +A E
 Sbjct: 1070 EQEVAELKQKILLFGCEKEEMNKETITWKEEGVKQDITLNLQEQKQKSAHVNSLAQDE 1129

Query: 176 RSLNLYRDKYQSSLSNIELLECQVKMLQGLGGI--MGQEPENKGDHDKSVRIYTSPCMIQ 233
 L + +K + L N L E LQ +L + + +E + K ++ T+ Q

Sbjct: 1130 TKLKAHLEKLEVDL-NKSLKENT--FLQEQVLKMLAEEDKRVSELTSLKKTDEEFQ 1186

Query: 234 E---HQETQKRLSEVWQKVSQQDDLIQELRNKL--AC--SNALVLEREKALIKLQADFA 285
H+++ K L + K + L +EL +L C + AL+ + LI + +

Sbjct: 1187 SLKSSHEKSNSKLED---KSLEFKLSEELAIQLDICCKTEALLEAKTNELINISSSKT 1243

Query: 286 SCTATH-RYPPSSSEECEDIKKILKHLQEOKDSQCLHVEEYQNLVKDLRVELEAVSEQKR 344
+ + + + + I + ++Q + E QN + + E+K

Sbjct: 1244 NAILSRISHCQHRRTTKVKEALLIKTCTVSELEAQLRQLTEEQNTLNISFQQATHQLEEKE 1303

Query: 345 NIMKDMKLELD-LHGLREETSABIERKDKDITILQCRLQELQLEFTET-QKLTLLKKDKF 402
N +K M K +++ L +E + + + + +L+ E +E +TL K++

Sbjct: 1304 NQKSM-KADIESLVTEKEALQKEGNGQQAASEKESCITQLKKELSENINAVTLMKEE- 1361

Query: 403 LQEKDEMLQELEKLTQVQNSLLKKEKELEKQQCMATELEMTVKEAKQDKSKEAECKALQ 462
L+EK + L K+LT + N L+ L +++ + L E K + . + + L

Sbjct: 1362 LKEKKVEISSLSKQLTDL-NVQLQNSISLSEKAAISSLRKQYDEEKCELLDQVQ--DLS 1418

Query: 463 AEVQKLKNSLEEAKQQLERLAAQQAACKEEAALAGCHLEDTQRKLQKGLLLDKQKA---- 518
+V L A +Q + + ++ K++A ++T ++LQ L L ++A

Sbjct: 1419 FKVDTLSEKISALEQVDDWSNKFSEWKKKAQSRFTQHONTVKELQIQLELKSKEAYEKD 1478

Query: 519 DTIQELQRELQMLQKESMAEKEQTSNRKRVEELSLELSEALRKLENSDKEKRQLQKTVA 578
+ I L+ EL K + E ++ ++E+ L +L++ +L+ +

Sbjct: 1479 EQINLLKEELDQONKRFDCCLKGEMEDDKSMEKKESNLET---ELKSQTARIMELEDHIT 1535

Query: 579 EQDMKMNDMLDRIKHQHQREQSGSIKCK-LEEDLQEATKLEDKREQLKKSKEHEKLMGEL 637
++ +++ + + +K+ + +Q I+ K L + LQ +L E+K ++K+++E +E ++

Sbjct: 1536 QKTIEIESLNEVLKN-YNQKQDIEHKELVQKLQHFQELGEEKDNRVKEAEKILTLENQV 1594

Query: 638 EALRQEFKKDKTKLENSRKLEENENLRAELQCCSTQLES-SLNKYNTSQVQIQDLNKE 696
+++ E + K K L+ + ++ + E L+A L+ +LES S K ++ + ++

Sbjct: 1595 YSMKAELETKKKELEHVNLVSVKSEELKA-LE---DRLESESAAKL---AELKRKAEQK 1647

Query: 697 IALQKESLMSLQAQLDKALQKEKHYLQTTITKEAYDALSRKSAACQDDLTQALEKLNHVT 756
IA K+ L+S Q++ +KE+ Y + T + L+ K + ++ EKL V

Sbjct: 1648 IAAIKKQLLS---QME---EKEEQYKKG--ESHLSLNTKLQEREREVHILEEKLKSVE 1699

Query: 757 S---ET---KSLQSLTQTQEKKAQLEEEII-AYEERMKKLNTLRLRGFHOESELEV 808
S ET +S + T++++A + + YEE++ L L E E +

Sbjct: 1700 SSQSETLIVPRSAKNVAAYTEQEEADSGQCVQKTYEEKISVLQRNLT-----EKEKLL 1752

Query: 809 HAFDKKLEEMSCQVLQWQKHQNDLKMLAAKEEQREFQEEMAALKENLLEDDKEPCCLP 868
++ EE + + Q+Q L L E + E Q + L+E L+K+ +

Sbjct: 1753 QRVGQKEKEETVSSHFMRCQYQERLIKLEHAEAKQHEDQSMIGHLQEELKKNKYSLIV 1812

Query: 869 QWSVPKDTCLRYRGNDQIMTNLEQ-WAKQKQVANKE-LGNQLREQ-VNYIAKLSGEKDHL 925
V K+ + N Q NLE + QK EK L Q+ EQ + + + +

Sbjct: 1813 AQHVEKEGGK---NNIQAKQNLNVFDDVQKTLQEKELTCQILEQKIKELDSCLVRQKEV 1869

Query: 926 HSV-MVHLQQENKKL 940
H V M L + +KL+

Sbjct: 1870 HRVEMEELTSKYEKLQ 1885

Score = 227 (34.1 bits), Expect = 2.5e-14, P = 2.5e-14
Identities = 160/716 (22%), Positives = 318/716 (44%)

Query: 233 QEHQETQKRLSEVWQKVSQQDDLIQE-LRNKLACSNAIV-LEREKALIKL-QADFASCTA 289
+E +TQ ++ +V + L + ++ L S++ L R + L + D S TA

Sbjct: 53 RESGDTQSFAKLQLRVPSVESLFRSPIKESLFRSSSKESLVRTSSRESLNRLDLDSSTA 112

Query: 290 THRYPPSSSEECEDIKKILKHLQEOKDSQCLHVEEYQNLVKDLRVELEAVSEQKRNMKD 349
+ P E ED+ L +++ Q L + + R + + + + ++

Sbjct: 113 SFDPPSDMDSEADLVGNSDSLNEQLIQRLR--RMERSLSSYRGKYSVLTAYQMLQRE 170

Query: 350 MMKLELDLHGLREETSABIERKDKDIT-ILQCRLQELQLEFTETQKLTLLKKDKFLQEKDE 408
KL+ G+ ++ +DK + I + R +ELQ++ + L + D L+EKD+

Sbjct: 171 KKKLQ----GILSQS-----QDKSLRRIAELR-EELQMDQQAQKHLQEEDASLEEKDQ 219

Query: 409 MLQELEKLTQVQNSLLKKEKELEKQQCMATELEMTVKEAKQDKSKEAECKALQAE---V 465
+ L+ +++ ++ L ++ + + +LE + + +++ E++ + + + V

Sbjct: 220 YISVLQTVQVSLKQLRLNGPMNVDLKPLF-QLEPQAEVFTKEENPESDGEFPVVEDGTSV 278

Query: 466 QKLKNSLEEAKQQLERL--AQQAAC-KEEAALAGCHLEDTQRKLQKGLL-LDKQKADTI 521
+ L+ + K+Q L ++ Q KE+ L E Q +L+ L L+K K +

Sbjct: 279 KTLLETQQRVQRQENLLKRCKETIQSHKEQCTILTSEKEALQEQLDERLQELEKIKDLHM 338

Query: 522 QELQRELQMLQKESMAEKEQTSNRKRVEELSLELSEALRKLENSDKEKRQLQKTVAEQD 581
E + + + L+ ++ E+ + + E ++ EL E + R K + Q

Sbjct: 339 AEXTKLTQLRDAKNLIEQLEQDKGMVIAETKRMHETLEMKEEIIAQLRSRIKQMTTQG 398

Query: 582 MKMNDMLDRIKHQHREQGSIKCKLEEDLQEAT-KLEDKREQLK---KSKEHEKL-MEGE 636
 ++ + ++ + E+ + +EA KL + EQ+K K+ E E++ ++ E
 Sbjct: 399 EELREQKEKSERAAFELEKALSTAQKTEEARRKLKAEMDEQIKTIEKTSEERISLQQE 458

Query: 637 LEALRQEFKK-KDKTLKENSRLKEEENENLRAELQCCSTQLESSLNKYNTSQQVIQDLNK 695
 L ++QE K+ +E KL++ +E EL +L L T ++ Q+ K
 Sbjct: 459 LSRVKQEVVDVMKKSSEEQIAKLQKLHEK---ELARKEQELTKKLQ---TREREFQEQMK 512

Query: 696 EIALQKESLSLQAQLDKALQKEKHYLQTTITKEAYDALSRKSAACQDDLTQALEKLN-H 754
 +AL+K L+ +K Q+ + + K+A S DL Q E
 Sbjct: 513 -VALEKSQSEYLLKISQEKEQQESLAELELQKKAILTESENKLR---DLQQAETYRTR 568

Query: 755 VTSETKSLQOQSLTQTQEKKAQLEEEIIAYEERMKKLNTELRLRGFHQSELEV--HAFD 812
 + SL++SL QE K Q ++ + E K N E+ + H+ +ELE H D
 Sbjct: 569 ILELESSLEKSL---QENKNQSKDLAVHLEAEKNKHNKEITVMVEKHK-TELESKHHQD 624

Query: 813 KKLEEMSCQVLQWQKHQNDLKMLAAKEEQLRE-----FQEMAALKENLLED-DK 862
 E QVL+ +Q+Q +++ L K EQ +E FQ + + E LE D
 Sbjct: 625 ALWTE-KLQVLK--QQYQTEMEKLEKCEQEKETLLKDKKEIIFQAHIEEMNEKTLEKLDV 681

Query: 863 EPCCLPQWSVPKDTCLRYRGNDQIMTNLEQWAKQKQVANEKLGNLREQVNYIAKLSGEK 922
 + L S + + + ++ L Q ++L ++ EQ N+ +
 Sbjct: 682 KQTELE--SLSELSEVLKARHKLEELSVLKDQTDKMKQELEAKMDEQKNHHQQQVDSI 739

Query: 923 DHLHSMVHLQQENKKLKEIEEKKM 948
 H V + Q+ K LK +I + ++
 Sbjct: 740 IKEHEVSI--QRTEKALKDQINQLEL 763

Score = 183 (27.5 bits), Expect = 1.3e-09, P = 1.3e-09
 Identities = 132/584 (22%), Positives = 251/584 (42%)

Query: 409 MLQELEKLTQVQNSLLKKEKELEKQCCMATELEMTVKEAK-QDKSKEAECKALQAEVQK 467
 M ++L++K+++ Q L + + +T M + + ++ E + Q
 Sbjct: 1 MFKKLLKQKISEEQQLQALAPAQASSNSSTPTRMRSRTSSFTEQLDEGTPNREGSDTQS 60

Query: 468 LKNSLE-EAKQQERLAAQQAQCKEEAALAGCHLEDTRKQLQKGLLLDKQKA--DTIQEL 524
 L+ E L + ++ + +R+ L LD A D ++
 Sbjct: 61 FAQKLQLRVPSVESLFRSPIKESLFRSSSKESLVRTSSRESLNRLDLSSTASFDPPSDM 120

Query: 525 QRELQMLQKESMAEKEQTSNRKRVEELSL-----ELSEALRKLNSDKERQLQKTVAE 579
 E + L S KEQ R R E SL + SE + + +EK++LQ +++
 Sbjct: 121 DSEAEDLVGNSDSLNEQLIQLRLRMERSLSSYRGKSELVTAYQMLQREKKKLGILSQ 180

Query: 580 -QDMKMDMLDRIKHQHREQGSIKCKLEE---DLQEATK---LLEDKREQLKKSKEHEKL 632
 QD + + + + +Q + K EE L+E + +L+ + LK+ + +
 Sbjct: 181 SQDKSLRRIAREELQMDQQAQKHLQEEFDASLEEKDQYISVLQTQVSLKQRLRNGPM 240

Query: 633 MEGELEALRQ-EFKKKDKTLKENSRLKEE---ENENLRAELQCCSTQLESSLNKYNTSQ 688
 L+ L Q E + + T +EN E E+ L+ +++ N ++
 Sbjct: 241 NVDVLKPLPQLEPQAEVFTKEENPESDGPVVEDGTSVKTLLETLQQRVKRQENLLKRCKE 300

Query: 689 VIQDLNKEIALQKESLSLQAQLDKALQKEKHYLQTTITKEAYDALSRKSAACQDDLTQA 748
 IQ ++ L +LQ QLD+ LQ E ++ E +++ A +L +
 Sbjct: 301 TIQSHKEQCTLLTSEKEALQEQLDERLQ-ELEKIKDLHMAEKTCLITQLRDA--KNLIEQ 357

Query: 749 LEK-LNHVTSETKSLQOQSLTQTQEKKAQLEEEIIAYEERMKKLNTELRLRGFHQSELE 807
 LE+ V +ETK + + +T E K EEEI R+K++ T+ +LR Q+ + E
 Sbjct: 358 LEQDKGMVIAETK---RQMHTLEMK---EEEIAQLRSRIKQMTTQGEELR--EQKEKSE 409

Query: 808 VHAFDKKLEEMSCQVLQWQKHQNDLKMLAAKEEQLREFQ----EEMAALKENLLEDDKE 863
 AF EE+ + QK + K+ A +EQ++ + EE +L++ L +E
 Sbjct: 410 RAAF---EELEKALSTAQKTEEARRKLKAEMDEQIKTIEKTSEERISLQQELSRVKQE 465

Query: 864 PCCLPQWSVPKDTCLRYRGNDQIMTNLEQ-WAKQKQVANEKLGNLRL-----EQVNYIAK 917
 + + S + +L + +++ + EQ K+ + + Q++ Q Y+ K
 Sbjct: 466 VVDVMKKSSEEQIAKLQKLHEKELARKEQELTKKLQTREREFQEQMKVALEKSQSEYL-K 524

Query: 918 LSGEKDLHSMVHL-LQENKKLKEIEEK---KMKAEATRLCTKALGPSRTESTQREK 972
 +S EK+ S+ + L+ + K+ + E E K + +AE R L S +S Q K
 Sbjct: 525 ISQEKEQQESLAELELQKKAILTESENKLRDLQQAETYRTRILELESSLEKSLQENK 584

Pedant information for DKFZphtes3_lg13, frame 1

Report for DKFZphtes3_lg13.1

{LENGTH} 1007
 {MW} 117480.77
 {PI} 5.90

[HOMOL] TREMBL:AF092090_1 product: "cp151"; Rattus norvegicus cp151 mRNA, partial cds.
0.0
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDL058w] 5e-15
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL058w] 5e-15
[FUNCAT] 09.10 nuclear biogenesis [S. cerevisiae, YDR356w] 1e-11
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YDR356w] 1e-11
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YDR356w] 1e-11
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKR095w] 1e-08
[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YKR095w] 1e-08
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YLR309c] 1e-08
[FUNCAT] 1 genome replication, transcription, recombination and repair [M. jannaschii, MJ1322] 4e-06
[FUNCAT] 09.13 biogenesis of chromosome structure [S. cerevisiae, YLR086w] 9e-06
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-04
[FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-04
[FUNCAT] 03.25 cytokinesis [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-04
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YJR134c] 5e-04
[EC] 3.6.1.32 Myosin ATPase 1e-16
[PIRKW] nucleus 3e-10
[PIRKW] phosphotransferase 6e-09
[PIRKW] duplication 2e-06
[PIRKW] citrulline 2e-12
[PIRKW] tandem repeat 1e-16
[PIRKW] endocytosis 2e-13
[PIRKW] heart 8e-13
[PIRKW] transmembrane protein 1e-13
[PIRKW] serine/threonine-specific protein kinase 6e-09
[PIRKW] zinc finger 2e-13
[PIRKW] metal binding 2e-13
[PIRKW] DNA binding 4e-12
[PIRKW] muscle contraction 1e-16
[PIRKW] acetylated amino end 1e-11
[PIRKW] actin binding 1e-16
[PIRKW] mitosis 5e-15
[PIRKW] microtubule binding 5e-15
[PIRKW] ATP 1e-16
[PIRKW] thick filament 1e-16
[PIRKW] phosphoprotein 4e-16
[PIRKW] skeletal muscle 2e-14
[PIRKW] calcium binding 2e-12
[PIRKW] alternative splicing 1e-16
[PIRKW] coiled coil 1e-16
[PIRKW] P-loop 1e-16
[PIRKW] heptad repeat 3e-10
[PIRKW] methylated amino acid 1e-16
[PIRKW] immunoglobulin receptor 2e-06
[PIRKW] peripheral membrane protein 2e-13
[PIRKW] cardiac muscle 8e-13
[PIRKW] hydrolase 1e-16
[PIRKW] microtubule 3e-10
[PIRKW] muscle 8e-13
[PIRKW] EF hand 2e-12
[PIRKW] cytoskeleton 2e-15
[PIRKW] hair 2e-12
[PIRKW] calmodulin binding 2e-13
[PIRKW] Golgi apparatus 3e-10
[PIRKW] myosin heavy chain 1e-16
[PIRKW] conserved hypothetical P115 protein 1e-07
[PIRKW] centromere protein E 5e-15
[PIRKW] unassigned Ser/Thr or Tyr-specific protein kinases 6e-09
[PIRKW] calmodulin repeat homology 2e-12
[PIRKW] myosin motor domain homology 1e-16
[PIRKW] alpha-actinin actin-binding domain homology 2e-07
[PIRKW] plectin 2e-07
[PIRKW] trichohyalin 2e-12
[PIRKW] pleckstrin repeat homology 8e-08
[PIRKW] ribosomal protein S10 homology 2e-07
[PIRKW] giantin 3e-13
[PIRKW] protein kinase homology 6e-09
[PIRKW] protein kinase C zinc-binding repeat homology 8e-08
[PIRKW] kinesin motor domain homology 5e-15
[PIRKW] human early endosome antigen 1 2e-13
[PIRKW] M5 protein 1e-07
[PROSITE] LEUCINE ZIPPER 7
[PROSITE] MYRISTYL 2
[PROSITE] CAMP_PHOSPHO_SITE 2
[PROSITE] CK2_PHOSPHO_SITE 20

[illegible]

```

COILS .....
SEQ   NEKLGNQLREQVNYIAKLSGEKDLHLSVMVHLQQENKKLKEIEEKKMKAENTRLCTKAL
SEG   .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD   hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....

SEQ   GPSRTESTQREKVCGLGWKGLPQDMGQRMDLTKYIGMPHCPGSSYC
SEG   .....
PRD   cchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....

```

Prosites for DKFZphtes3_lg13.1

PS00001	52->56	ASN_GLYCOSYLATION	PDOC00001
PS00001	684->688	ASN_GLYCOSYLATION	PDOC00001
PS00004	240->244	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	415->419	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	74->77	PKC_PHOSPHO_SITE	PDOC00005
PS00005	110->113	PKC_PHOSPHO_SITE	PDOC00005
PS00005	238->241	PKC_PHOSPHO_SITE	PDOC00005
PS00005	290->293	PKC_PHOSPHO_SITE	PDOC00005
PS00005	392->395	PKC_PHOSPHO_SITE	PDOC00005
PS00005	396->399	PKC_PHOSPHO_SITE	PDOC00005
PS00005	444->447	PKC_PHOSPHO_SITE	PDOC00005
PS00005	503->506	PKC_PHOSPHO_SITE	PDOC00005
PS00005	544->547	PKC_PHOSPHO_SITE	PDOC00005
PS00005	566->569	PKC_PHOSPHO_SITE	PDOC00005
PS00005	600->603	PKC_PHOSPHO_SITE	PDOC00005
PS00005	650->653	PKC_PHOSPHO_SITE	PDOC00005
PS00005	655->658	PKC_PHOSPHO_SITE	PDOC00005
PS00005	735->738	PKC_PHOSPHO_SITE	PDOC00005
PS00005	876->879	PKC_PHOSPHO_SITE	PDOC00005
PS00005	968->971	PKC_PHOSPHO_SITE	PDOC00005
PS00006	39->43	CK2_PHOSPHO_SITE	PDOC00006
PS00006	53->57	CK2_PHOSPHO_SITE	PDOC00006
PS00006	68->72	CK2_PHOSPHO_SITE	PDOC00006
PS00006	116->120	CK2_PHOSPHO_SITE	PDOC00006
PS00006	190->194	CK2_PHOSPHO_SITE	PDOC00006
PS00006	250->254	CK2_PHOSPHO_SITE	PDOC00006
PS00006	296->300	CK2_PHOSPHO_SITE	PDOC00006
PS00006	439->443	CK2_PHOSPHO_SITE	PDOC00006
PS00006	444->448	CK2_PHOSPHO_SITE	PDOC00006
PS00006	471->475	CK2_PHOSPHO_SITE	PDOC00006
PS00006	520->524	CK2_PHOSPHO_SITE	PDOC00006
PS00006	536->540	CK2_PHOSPHO_SITE	PDOC00006
PS00006	566->570	CK2_PHOSPHO_SITE	PDOC00006
PS00006	576->580	CK2_PHOSPHO_SITE	PDOC00006
PS00006	650->654	CK2_PHOSPHO_SITE	PDOC00006
PS00006	674->678	CK2_PHOSPHO_SITE	PDOC00006
PS00006	804->808	CK2_PHOSPHO_SITE	PDOC00006
PS00006	888->892	CK2_PHOSPHO_SITE	PDOC00006
PS00006	963->967	CK2_PHOSPHO_SITE	PDOC00006
PS00006	968->972	CK2_PHOSPHO_SITE	PDOC00006
PS00007	135->143	TYR_PHOSPHO_SITE	PDOC00007
PS00008	207->213	MYRISTYL	PDOC00008
PS00008	599->605	MYRISTYL	PDOC00008
PS00029	83->105	LEUCINE_ZIPPER	PDOC00029
PS00029	90->112	LEUCINE_ZIPPER	PDOC00029
PS00029	97->119	LEUCINE_ZIPPER	PDOC00029
PS00029	104->126	LEUCINE_ZIPPER	PDOC00029
PS00029	403->425	LEUCINE_ZIPPER	PDOC00029
PS00029	410->432	LEUCINE_ZIPPER	PDOC00029
PS00029	918->940	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphtes3_lg13.1)

DKFZphtes3_1k11

group: cell structure and motility

DKFZphtes3_1k11 encodes a novel 589 amino acid protein with strong similarity to *Mus musculus* actin-binding protein (ENC-1).

Ectoderm-neural cortex-1 protein (ENC-1) is an early and highly specific marker of neural induction in vertebrates. The protein is related to the kelch family proteins and is expressed during early gastrulation in the prospective neuroectodermal region of the epiblast and later in development throughout the nervous system (NS). ENC-1 functions as an actin-binding protein organising the actin cytoskeleton during neural differentiation and development of the NS. The novel protein is highly similar to ENC-1.

The new protein can find application in modulation of cyto skeleton organisation in human testicular cells.

strong similarity to mouse ENC-1

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 3525 bp

Poly A stretch at pos. 3515, polyadenylation signal at pos. 3499

```

1 GGTGGAGAGC CGGCCGACGG GAGCCGCGGC GGAGCCTGTT GAGCTCGCGC
51 GGGCTGCCGG GAGTGGTCTC TGAGGCGGCG GCGGCGGCGG GGATCGTCTC
101 CGGCACTGGC GCACCATGTC GGTCACTGTC CATGAGACCC GCAAGTCGCG
151 GAGCAGCAGC GGGTCCATGA ACGTCACCCCT CTTCCACAAG GCCTCCACCC
201 CGGACTGTGT GCTGGCCAC CTCAACACGC TTCGCAAGCA CTGCATGTTT
251 ACCGACGTCA CACTCTGGGC GGGCGACCGT GCCTTCCCTT GTCAACCGTGC
301 CGTGCTGGCC GCCTCTAGCC GCTATTTTGA GGCCATGTTT AGCCATGGCC
351 TTCGGGAGAG CCGGGATGAC ACTGTCAACT TCCAGGACAA CCTGCACCCG
401 GAGGTGCTGG AGCTGCTGCT GGACTTTGCC TACTCCTCAC GCATCGCCAT
451 CAACGAGGAG AACGCTGAGT CACTGCTGGA GGCAGGCGAC ATGCTGCAGT
501 TCCACGATGT GCGGGATGCT GCCGCCGAGT TCCTGGAGAA GAACCTTTTC
551 CCCTCCAAC TCCCTGGCAT GATGCTGCTC TCGGACGCC ACCAGTGCCG
601 CCGGCTGTAT GAGTTCTCCT GCGCATGTG CCTGGTGCAC TTTGAGACGG
651 TGAGGAGAGC CGAGGACTTC AACAGCCTGT CCAAGGACAC ACTGCTGGAC
701 CTATCTCGA GTGATGAGCT GGAGACCGAG GACGAGCGGG TGGTCTTCGA
751 GGCCATCCTC CAGTGGGTGA AGCAGCACCT GGAGCCACGG AAGGTCCACT
801 TGCCCGAGCT CCTCCGAGC GTGCGTCTGG CCTTGCTGCC GTCCGACTGC
851 CTCGAGGAGC CCGTCTCCAG CGAGGCCCTC CTATGGCAG ACAGAGCGAC
901 CAAGCTTATC ATGGATGAGG CCCTGCGCTG CAAGACCAGG ATCCTGCAGA
951 ATGATGGCGT GGTCAACAGC CCCTGTGCCC GGCCACGCAA GGGGGGCCAC
1001 ACCTACTCTA TCCTGGGGG CCAGACCTTC ATGTGTGACA AGATCTACCA
1051 GGTGGACCAC AAGGCCAAGG AGATCATCCC CAAGGCCGAC CTGCCAGCC
1101 CCCGGAAGGA GTTCAGCGCC TCAGCGATCG GCTGCAAGGT CTATGTGACG
1151 GGGGGCAGGG GCTCCGAGAA CGGGGTCTCC AAGGATGTCT GGGGTGACGA
1201 CACCGTATAT GAGGAATGGT CCAAGGCGGC GCCCATGCTG ATTGCCCGCT
1251 TTGGCCATGG CTCAGCTGAG CTGGAGAACT GCCTCTATGT GGTGGGGGGA
1301 CACACATCCC TGGCAGGGGT CTTCCCGGCC TCGCCTTCTG TCTCCCTGAA
1351 ACAAGTGGAG AAATACGACC CTGGGGCCAA CAAGTGGATG ATGGTGGCCC
1401 CCTTGGCGGA TGGCGTCAGC AATGCCGAG TGGTGAGTGC CAAGCTGAAG
1451 CTCTTTGTTT TCGGAGGAAC CAGCATCCAC CGGGACATGG TGTCCAAGGT
1501 CCAGTGCTAT GACCCCTCGG AGAACAGGTG GACGATCAAG GCCGAGTGCC
1551 CCCAGCCTTG GCGGTACACA GCGGCTGCCG TCCTGGGCAG CCAGATCTTC
1601 ATCATGGGAG GTGACACGGA ATTACAGGCC GCCTCGGCCT ACCGCTTTGA
1651 CTGTGAGACC AACCAAGTGA CGCGGATTGG GGACATGACT GCCAAGCGCA
1701 TGTCTTGCCA TGCCCTGGCT TCCGGCAACA AGCTCTATGT GGTGGGGGGC
1751 TACTTTGGGA CCCAGAGGTG TAAGACTCTG GACTGCTATG ACCCCACTTC
1801 AGATACATGG AACTGCATCA CCACAGTGCC CTACTCACTT ATCCCCACGG
1851 CCTTTGTGAG CACCTGGAAG CACCTGCCCG CGTGAGGAGC ACCTGCTGAG
1901 CCCAGCCAGA CCGCGGCCTT CAGTGTACAC GCGTGGCCTT GCTTGTCTGC
1951 CACAGCGGGA GCTAAGCCGG CCCTGGGCCA GCACCTCCGAG AGGTGGAAGG
2001 GGGCCTGCCA GCTCTGGGGA GCAGCAGCCT TGGGCTGTTC TGAGCTTTAG
2051 GCAAGAGAAG AGAAGCATCT CTTGCATCCG TGCCCTGGG GGCCTCTTCA
2101 GCTTTGCGAT GGTGTGTTGG AAGACATACC TCCAGAGGG GCATGGACTG
2151 CCACAGGAC TGACCTTGGC GTCGGGGAGA AGGACACTTG CAGAGCCTTG
2201 AGATCACTTG TTTGGCAGGT CCTGGACTGG GGGCGGGCAG GCAGGGGCGC
2251 GAGGGCGCCC CGGGTGGGCT TTGGGGCTGC GGCACTGCCA CACATCCTTT
2301 CCCTCCTGGC CTGCCCTGCT GGGGCTCTAC TGCCATCTAT AGATGGTGTG
2351 CTGGGCTTGG GAAACTAGGT TCCAGGGGGT TGAGACCAGA AAGGTGACCA
2401 AGACAGATTT TTTAAGGTGC AGAAACTGCA GGGGGGCTC AGTGACATCC
2451 ATGAGGCCTT ATTAGCAAAG GACACCCAGA CCTCAAGGT TGTGGGGCCC
2501 CTCCACAAA GCTGTAAGTC CCAGCCACCC TACTCAGGGC CTGTCTCAGT
2551 GCTGTGGGCC GGTGGGGACA CAGTTGCTCG TGGCCACTCA GTGGAGCTGC
2601 GCCTGCAGCA GACTCAAGGC TCCGAGTGCC CTGGGGGTCA CCCTCCCTTC
2651 CCCCTCCTCA GAGCCACCC TGAGAGGCAG CAGTGACCCC CATGGCACAC
2701 ACCTGCCAAC AGCACTGGGG GCTTCTCCCC AGGAGACCAC GCTGCCCTTC

```



```

2751 AAGACCAGGA GCAGCTGTGA GCTGGAGACA GCAGAGGGAC CCCAGGGTGT
2801 CCCCTGCAGA TCCCACCAGG GCCGCATCCA TCTCAGTGTG GAGGACAGTG
2851 ACGGGACCCT CACCATCCTC TTGCGTTTGT GCCCCCATTT GCTCCCTGAG
2901 CTCCAAGATA AGAATGGCCC CGAGAGAACT GCTGAACATT TGTTCATTGC
2951 TGTCACTCCG TGAGTCACTG GGGTCCCTCA CCAGCACCTC CCTGACACCT
3001 GGGCTATGGA GAGGTGGCGG CCTGTCACTG ACCATCCTAA TGCCTCTCGC
3051 TCACTCCCAA GCCACCATTG GAGAGGGAGG GGTGTTGGTG CCCTGACAGG
3101 GACTGGGCAG GGTGTCCAAA CTTGGGGCTT CCCAGGCACC TGCAGTGTGA
3151 ACACTGCTTG GCTGGCTCAA GATTAGGGCC GCGGAGGGGG CTGTGCACAT
3201 ACCAGTTACT TAAGCAGCCA CGAGTGTCCC CCATGCCTTG GTGCGGGTCC
3251 TGGAGGCTTC TTGGGGGTGG GACCTTTGGG CAGGGTTTGC CCACTGACGC
3301 GCCCGCCATG GGGCACTGGC TGCATGGGGC TCCTTGGACC CTGTAGACCC
3351 AGCAGGAGCC TGGCCGCGGG GACTGCAGGG AGGGTGCGTG GACCCGTGGG
3401 GTTGCTTCAT TGAGATAAAG CACACTTATC ACATAGCACA AAGGACGTGC
3451 CATGGTGCTT TCCCCAAAAG TTGTGTTGCT TTTATCAGTT TTCTAACTTA
3501 ATAAAAAGAG TTGAGAAAAA AAAAA

```

BLAST Results

No BLAST result

Medline entries

98350113:

Cloning of human ENC-1 and evaluation of its expression and regulation in nervous system tumors.

97252647:

ENC-1: a novel mammalian kelch-related gene specifically expressed in the nervous system
encodes an actin-binding protein.

98234394:

NRP/B, a novel nuclear matrix protein, associates with p110(RB) and is involved in neuronal differentiati

Peptide information for frame 2

ORF from 116 bp to 1882 bp; peptide length: 589

Category: strong similarity to known protein

Classification: Cell structure/motility

```

1 MSVSVHETRK SRSSTGSMNV TLFHKASHPD CVLAHLNLTNR KHCMTDVTLL
51 WAGDRAFPCH RAVLAASSRY FEAMFSLGLR ESRDDTVNFQ DNLHPEVLEL
101 LLDFAVSSRI AINEENAESL LEAGDMLQFH DVRDAAAEFL EKNLFPSNCL
151 GMLLSDAHQ CRRLYEFSWR MCLVHFETVR QSEDFNSLSK DTLLDLISSD
201 ELETEDERVV FEAILQWVKH DLEPRKVHLP ELLRSVRLAL LPSDCLQEAV
251 SSEALLMADE RTKLIMDEAL RCKTRILQND GVVTSPCARP RKAGHTLLIL
301 GGTFTMCDKI YQVDHKAKEI IPKADLPSPR KEFSASAIGC KVVVTGGRGS
351 ENGVS KD VVW YDTVHEEWSK AAPMLIARFG HGSAELENCL YVVGHTSLA
401 GVFPASPSVS LKQVEKYDPG ANKWMVAPL RDGVSNAAVV SAKLKL FVFG
451 GTSIHRMVS KVQCYPDSEN RWTIKAECPO PWRYTAAAVL GSQIFIMGGD
501 TEFTAASAYR FDCETNQWTR IGDMTAKRMS CHALASGNKL YVVGGYFGTQ
551 RCKTLDCYDP TSDTWCNITT VPYSLIPTAF VSTWKHLPA

```

BLASTP hits

Entry MMU65079_1 from database TREMBL:

gene: "ENC-1"; product: "actin-binding protein"; Mus musculus
actin-binding protein (ENC-1) mRNA, complete cds.

Score = 2402, P = 1.9e-249, identities = 440/589, positives = 513/589

Entry AF059611_1 from database TREMBLNEW:

gene: "NRPB"; product: "nuclear matrix protein NRP/B"; Homo sapiens
nuclear matrix protein NRP/B (NRPB) mRNA, complete cds.

Score = 2400, P = 3.0e-249, identities = 440/589, positives = 512/589

Entry AF010314_1 from database TREMBL:

gene: "PIG10"; product: "Pig10"; Homo sapiens Pig10 (PIG10) mRNA,
complete cds.

Score = 1745, P = 7.8e-180, identities = 335/507, positives = 403/507

Alert BLASTP hits for DKFZphtes3 1k11, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphtes3 1k11, frame 2

Report for DKFZphtes3 1k11.2

```
[LENGTH]      589
[MW]           65923.45
[pI]           6.10
[HOMOL]        TREMBL:MMU65079_1 gene: "ENC-1"; product: "actin-binding protein"; Mus musculus
actin-binding protein (ENC-1) mRNA, complete cds. 0.0
[FUNCAT]       10.05.99 other pheromone response activities [S. cerevisiae, YHR158c]
2e-09
[BLOCKS]       BL01016D Glycoprotease family proteins
[PIRKW]        zinc finger 1e-08
[PIRKW]        DNA binding 1e-08
[PIRKW]        transcription factor 1e-08
[SUPFAM]       POZ domain homology 3e-68
[SUPFAM]       vaccinia virus 59K HindIII-C protein 1e-15
[SUPFAM]       A55R protein 5e-29
[SUPFAM]       hypothetical protein YHR158c 4e-08
[SUPFAM]       A55R protein middle region homology 5e-29
[SUPFAM]       myxoma virus M9-R protein 1e-14
[SUPFAM]       A55R protein carboxyl-terminal homology 5e-29
[KW]           Alpha Beta
```

SEQ MSVSVHETRKSRSSTGSMNVTLFHKASHPDCVLAHLNLTLRKHCMTDVTWLWAGDRAFPCH
PRD cccccccccccccccccceeeeeccccchhhhhhhhhhhhhhhheeeeeccccchhhh

SEQ RAVLAASSRYFEAMFSHGLRESRDDTVNFDNLHPEVLELLLDFAFYSSRIATNEENAESL
PRD hccccccccccccccccchhhhhheeecccccchhhhhhhhhhhhhccceeehhhhhhhh

[illegible]

```
SEQ      QSEDFNSLSKDTLLDLISSELETEDERVVF EAILQVVKHDLPRKVHLPELLRSVRLAL
PRD      hhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhh
```

SEQ LPSDCLQEAVSSEALLMADERTKLIMDEALRCKTRIQLQNDGVVTSPCARPRKAGHTLLIL
PRD ccchhhhhhhhhhhhhccchhhhhhhhhhhhhhhhhhhccccccccccccccccceeeee

SEQ GGQTFMCDKIIYQVDHKAKEIIPKADLPSPRKEFSASAIGCKVYVTGGRGSENGVSKDVWV
PRD cccccccceeeeeccccccccccccccccccccccccceeeeeceeeeeccccccccccceeee

SEQ YDTVHEEWSKAAPMLIARFGHGSAALENCLYVVGGHSTLAGVFPA SPVS LKQVEKYDPG
PRD CC

SEQ ANKWMMPVAPLRDGVSNAAVVSAKLKLFVFGGTSIHRDMVSKVQCYPDSENRTIKAECPQ
PRD cccceeecccccccccccceeeccccceeeccccccccccccceeecccccccccccccccc

SEQ PWRYTAAAVLGSQIFIMGGDTEFTAASAYRFDCE TNQWTRIGDMTAKRMSCHALASGNKL
PRD CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ YVVGGYFGTQRCKTLDCYDPTSDTWN CITTVPYSLIPTAFVSTWKHLPA
PRD EEECC

(No Prosite data available for DKFZphtes3 1k11.2)

(No Pfam data available for DKFZphtes3 1k11.2)

DKFZphtes3_in3

 group: signal transduction

DKFZphtes3_in3 encodes a novel 1196 amino acid protein with similarity to *S. pombe* Tup1 protein.

The protein contains 1 WD-40 repeat, which is typical for the beta-transducin subunit of G-proteins. The beta subunits seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition. In addition, a RGD site is present.

The new protein can find application in modulating/blocking G-protein-dependent pathways.

similarity to Tup1p

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: /map="6q24"

Insert length: 5277 bp

Poly A stretch at pos. 5267, polyadenylation signal at pos. 5244

```

1 GCTGCATAAA GCTGAGAGAT GCCTACAGCT GAGAGTGAAG CAAAAGTAAA
51 AACCAAAGTT CGCTTTGAAA AATTGCTTAA GACCCACAGT GATCTAATGC
101 GTGAAAAGAA AAAACTGAAG AAAAACTTG TCAGGTCTGA AGAAAACATC
151 TCACCTGACA CTATTAGAAG CAATCTTCAC TATATGAAAG AAACATCAAG
201 TGATGATCCC GACACTATTA GAAGCAATCT TCCCATATT AAAGAAACTA
251 CAAGTGATGA TGTAAAGTGT GCTAACACTA ACAACCTGAA GAAGAGCAGC
301 AGAGTCACTA AAAACAAATT GAGGAACACA CAGTTAGCAA CTGAAAATCC
351 TAATGGTGAT GCTAGTGTAG AGGAAGACAA ACAAGGAAAG CCAAATAAAA
401 AGGTGATAAA GACGGTGCCC CAGTTGACTA CACAAGACCT GAAACCGGAA
451 ACTCCTGAGA ATAAGGTTGA TTCTACACAC CAGAAAACAC ATACAAAGCC
501 ACAGCCAGGC GTTGATCATC AGAAAAGTGA GAAGGCAAAAT GAGGGAAGAG
551 AAGAGACTGA TTTAGAAGAG GATGAAGAAT TGATGCAAGC ATATCAGTGC
601 CATGTAACTG AAGAAATGGC AAAGGAGATT AAGAGGAAA TAAGAAAGAA
651 ACTGAAAGAA CAGTTGACTT ACTTTCCCTC AGATACTTTA TTCCATGATG
701 ACAAACCTAG CAGTGAAAAA AGGAAAAAGA AAAAGGAAGT TCCAGTCTTC
751 TCTAAAGCTG AAACAAGTAC ATTGACCATC TCTGGTGACA CAGTTGAAGG
801 TGAACAAAAG AAAGAATCTT CAGTTAGATC AGTTTCTTCA GATTCTCATC
851 AAGATGATGA AATAAGCTCA ATGGAACAAA GCACAGAAGA CAGCATGCAA
901 GATGATACAA AACCTAAACC AAAAAAACA AAAAAAGAAG CTAAGCAGT
951 TGCAGATAAT AATGAAGATG TTGATGGTGA TGGTGTTTCA GAAATAACAA
1001 GCCGAGATAG CCGGTTTAT CCCAAATGTT TGCTTGATGA TGACCTTGTC
1051 TTGGGAGTTT ACATTCACCG AACTGATAGA CTTAAGTCAG ATTTTATGAT
1101 TTCTCACCCA ATGGTAAAAA TTCATGTGGT TGATGAGCAT ACTGGTCAAT
1151 ATGTCAAGAA AGATGATAGT GGACGGCCTG TTTTCATCTA CTATGAAAAA
1201 GAGAATGTGG ATTATATTCT TCCTATTATG ACCCAGCCAT ATGATTTTAA
1251 ACAGTTAAAA TCAAGACTTC CAGAGTGGGA AGAACAAATT GTATTTAATG
1301 AAAATTTTCC CTATTTGCTT CGAGGCTCTG ATGAGAGTCC TAAAGTCATC
1351 CTGTTCTTTG AGATTCTTGA TTTCTTAAGC GTGGATGAAA TTAAGAATAA
1401 TTCTGAGGTT CAAAACCAAG AATGTGGCTT TCGGAAAATT GCCTGGGCAT
1451 TTCTTAAGCT TCTGGGAGCC AATGGAATG CAAACATCAA CTCAAAACCT
1501 CGCTTGCAGC TATATTACCC ACCTACTAAG CCTCGATCCC CATTAAGTGT
1551 TGTTGAGGCA TTTGAATGGT GGTCAAAATG TCCAAGAAAT CATTACCCAT
1601 CAACACTGTA CGTAACTGTA AGAGGACTGA AAGTTCCAGA CTGTATAAAG
1651 CCATCTTACC GCTCTATGAT GGCTCTTCAG GAGGAAAAG GTAAACCAAGT
1701 GCATTGTGAA CGTCACCATG AGTCAAGCTC AGTAGACACA GAACCTGGAT
1751 TAGAAGAGTC AAAGGAAGTA ATAAAGTGGA AACGACTCCC TGGGCAGGCT
1801 TGCCGTATCC CAAACAAACA CCTCTTCTCA CTAATGCAG GAGAACGAGG
1851 ATGTTTTTGT CTTGATTCTT CCCACAATGG AAGAATATTA GCAGCAGCTT
1901 GTGCCAGCCG GGATGGATAT CCAATTATTT TATATGAAAT TCCTTCTGGA
1951 CGTTTCATGA GAGAATTGTG TGGCCACCTC AATATCATTT ATGATCTTTC
2001 CTGGTCAAAA GATGATCACT ACATCCTTAC TTCATCATCT GATGGCACTG
2051 CCAGGATATG GAAAAATGAA ATAAACAATA CAATACTTTT CAGAGTTTAA
2101 CCTCATCCTT CTTTGTGTTA CACGGCTAAA TTCCATCCAG CTGTAAGAGA
2151 GCTAGTAGTT ACAGGATGCT ATGATTCCAT GATACGGATA TGGAAAGTTG
2201 AGATGAGAGA AGATTCTGCC ATATTGGTCC GACAGTTTGA TGTTACAAAA
2251 AGTTTTATCA ACTCACTTTG TTTTGATACT GAAGGTCATC ATATGTATTC
2301 AGGAGATTGT ACAGGGGTGA TTGTTGTTTG GAATACCTAT GTCAAGATTA
2351 ATGATTGGA ACATTCAGTG CACCACTGGA CTATAAATAA GGAATTTAAA
2401 GAAACTGAGT TTAAGGGAAT TCCAATAAGT TATTGAGAGA TTCATCCCAA
2451 TGGAAAACGT TTGTTAATCC ATACCAAAGA CAGTACTTTG AGAATTATGG
2501 ATCTCCGGAT ATTAGTAGCA AGGAAGTTTG TAGGAGCAGC AAATTATCGG
2551 GAGAAGATTC ATAGTACTTT GACTCCATGT GGGACTTTTC TGTTTGCTGG
2601 AAGTGAGGAT GGTATAGTGT ATGTTTGGA CCCAGAAACA GGAGAACAAAG

```

```

2651 TAGCCATGTA TTCTGACTTG CCATTCAAGT CACCCATTTCG AGACATTTCCT
2701 TATCATCCAT TTGAAAATAT GGTGTCATTG TGTGCATTTC GGCAAAATGA
2751 GCCAATTCCT CTGTATATTT ACGATTTCCTA TGTGCCCAG CAGGAGGCTG
2801 AAATGTTCAA ACGCTACAAT GGAACATTTC CATTACCTGG AATACACCAA
2851 AGTCAAGATG CCTATGTAC CTGTCCAAAA CTACCCCATC AAGGCTCTTT
2901 TCAGATTGAT GAATTTGTCC ACGCTGAAAG TTCTTCAACG AAGATGCAGC
2951 TAGTAAACAA GAGGCTTGAA ACTGTCACAG AGGTGATACG TTCCTGTGCT
3001 GCAAAAGTCA AAAAAATCT CTCATTTACT TCACCACCAG CAGTTTCCTC
3051 ACAACAGTCT AAGTTAAAGC AGTCAAACAT GCTGACCGCT CAAGAGATTTC
3101 TACATCAGTT TGGTTTCACT CAGACCGGGA TTATCAGCAT AGAAAGAAAG
3151 CCTTGTAAAC ATCAGGTAGA TACAGCACC ACGGTAGTGG CTCCTTATGA
3201 CTACACAGCG AATCGATCAG ATGAACTAAC CATCCATCGC GGAGACATTA
3251 TCCGAGTGTT TTTCAAAGAT AATGAAGACT GGTGGTATGG CAGCATAGGA
3301 AAGGGACAGG AAGGTTATTT TCCAGCTAAT CATGTGGCTA GTGAAACACT
3351 GTATCAAGAA CTGCCTCCTG AGATAAAGGA GCGATCCCTT CCTTTAAGCC
3401 CTGAGGAAAA AACTAAAATA GAAAAATCTC CAGCTCCTCA AAAGCAATCA
3451 ATCAATAAGA ACAAGTCCCA GGAATTCAGA CTAGGCTCAG AATCTATGAC
3501 ACATTCAGAA ATGAGAAAAG AACAGAGCCA TGAGGACCAA GGACACATAA
3551 TGGATACACG GATGAGGAAG AACAAAGCAAG CAGGCAGAAA AGTCACTCTA
3601 ATAGAGTAAA GAATTGAAGA AAAGTTAAGA GCTGCCGAAA TGCACAGAGG
3651 TGAAAAATGAC AAACCAATG GAATTTCTCT TCAGAGTTCA GAATTTTCAG
3701 ATACTAAGGA GGAAGAAAGG ATCCACTACT TCTTGTCTT ATGAATGACT
3751 CTAGAAAAAT CAGAATCAAG TTGTGGGTGG AAAAAATCAAC GTGGCCTTTG
3801 AGTTCAGTTG TTATAAACCA TTGTGACTAT TGTGGTCAA AGTATTGGTA
3851 CTTATATTGT TAGTAATTGC ATCATAATTA CATTACCAGT GTTGAAAAAC
3901 TAATGAAGAA AACACTGTAA TTGCTACTCA GCAAAATGTA ATAAAAAGTG
3951 TTTGCGTTAT TAGGATGCTT GTTAAGTAAT CATTAAATAT TATTATATTG
4001 GTAATGGTTG TATGTGTGAT GCTATGCCCA GAATATGAAG TATCTGTTTT
4051 TGAAATTCAC TTTATTTAAA AGATAAGCAG CTGACTGGGC ACGGTGCCTC
4101 ATGCCTGTAA TCCTAGCACC TTGGGAGGCT GAGGCAGGTG GATCACCTAA
4151 GGTGAGGAGT TCAACAACAC CAGCCTGACC AACATGTGTA AACCCCATCT
4201 CTACTAAAAA TACAAAAATC AGCCGGGTCT CATGGCAGGC ACCTGTAATC
4251 CCATCTACTG AGGCAGGAGA ATTGCTTGAC CCAGGAGGCA GAGGTTGCAG
4301 TGAGCCAGA TCACGCCATT GCACTCCAGC CTGGGGGACA GAGCAAGACT
4351 CTATCTCCAA AAAACAAAAA AGATAAGCAG CTTTAGAATA TGGCGCATTC
4401 AAAACAGTCT CAGTAACAAA GACATTAAAA GAAAACAATT TACTTTCTAA
4451 TTAATAATTT GTGTTTCTTA AGATCAAATC ATATAGGTAA CTTTATAGAC
4501 CTAAATTAAT AGTGATTTT GGCTGGACTG GCAACAATGT TCCCAATGTC
4551 TTTACTTTTT AAAAAAGGCT TTTCATATTT AAGCACATAC CTATTTTGTA
4601 GACTTACATT GTTAAATATT TATTTTAAAT TTAATATTTT TACATTATTA
4651 TATTGCATTA TTTATTTTTT CTAAGTTCCA GAATAATAGT GTCATTATTA
4701 TAGACTATAT GTTTTGAAGT TTGATATTAT AATGGGATAT TCATTTTTTG
4751 TTCTTTTCTT GACTCCTTTC TCAAGTGTGT GATAAGGTCT GCTGATAAAA
4801 TATTTAACCC CAAGAAAGTG AAAACTAATA TAAATTAGA AAGACCTATC
4851 CAAATTAGAC AGTCAATTCC ATTAAATATA GAAGTGAGAA AAACAATGTT
4901 GGGCATTGAG GTGTAATTTT TGCCAGATG TATACCCAGT GTGAAATATC
4951 TTTTAATAAA AATATATTTG GCTCTTATCC CTGCACATGT AGAGGCATAA
5001 AAATTGGTAA ACATGTCCCG CTGTGTAGAA CTTTAAAAAA AAGGCATTTT
5051 TGAAAGTGTT GAGTGGCACT GATAACTGGT GAAGCCTACA GCCATCCGCC
5101 CAAAAGTCTG TTCTGATGGC ACTGAGTTTT CATTGTTCTG GATGTATAAG
5151 TCTGTGTGTC AGGTACAGCT GGGCCAGGCC AGCTTGAGTC ACTCTGTAC
5201 AAGCTTGTGT TTTTCTGTCT TGTGAATGCA CTTGATAATT TAAAAATAAA
5251 AATATCTGTT TCTCTGCAAA AAAAAAA

```

BLAST Results

Entry HS32B1 from database EMBL:
 Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 32B1
 Score = 4445, P = 0.0e+00, identities = 889/889

Entry U93816 from database EMBL:
 Human exon-trapped sequence from 6q24.
 Score = 965, P = 4.0e-35, identities = 193/193

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 19 bp to 3606 bp; peptide length: 1196
 Category: similarity to known protein

```

1 MPTAESEAKV KTKVRFEKLL KTHSDLMREK KKKKKKLVR EENISPDITR
51 SNLHYMKETT SDDPDITRSN LPHIKETTSO DVSAANTNNL KKSTRVTKNK
101 LRNTQLATEN PNGDASVEED KQKPNKKVI KTVPQLTTQD LKPETPENKV
151 DSTHQKTHTK PQPGVDHQKS EKANEGREET DLEEDDELMQ AYQCHVTEEM
201 AKEIKRRKIRK KLKEQLTYFP SDTLFHDDKL SSEKRRKKKE VPVFSKAETS
251 TLTISGDTVE GEQKKESSVR SVSSDSHQDD EISSMEQSTE DSMQDDTKPK
301 PKTKKKTKA VADNNEDVDG DGVHEITSRD SPVYPKCLLD DDLVLGVYIH
351 RTDRLKSDFM ISHPMVKIHV VDEHTGQYVK KDDSGRPVSS YYEKENVDYI
401 LPINTQPYDF KQLKSRLPEW EEQIVFNENF PYLLRGSDS PKVILFFEIL
451 DFLSVDEIKN NSEVQNQECG FRKIAWAFLK LLGANGNANI NSKLRLQLYY
501 PPTKPRSPLS VVEAFEWWSK CPRNHYPSTL YVTVRGLKVP DCIKPSYRSM
551 MALQEEKGKP VHCERHHESV SVDTEPGLEE SKEVIKWKRL PGQACRIPNK
601 HLFSLNAGER GCCLDFSHN GRILAAACAS RDGYPIILYE IPSGRFMREL
651 CGHLNIIYDL SWSKDDHYIL TSSSDGTARI WKNEINNTNT FRVLPHPSFV
701 YTAKFHFAVR ELVVTGCYDS MIRIWKVEMR EDSAILVRQF DVHKSFINSL
751 CDFTEGHHMY SGDCGTGIVV WNTYVKINDL EHSVHHWTIN KEIKETEFKG
801 IPTSYLEIHP NGKRLLIHTK DSTLRIMDLR ILVARKFVGA ANYREKIHST
851 LTPCGTFLFA GSEDGIVYVW NPETGEQVAM YSDLFFKSPI RDISYHPFEN
901 MVAFCAFQON EPILLYIYDF HVAQQAEMF KRYNGTFPLP GIHQSQDALC
951 TCPKLPHQGS FQIDFVHTE SSSTKMQLVK QRLETVTEVI RSCAAKVNKN
1001 LSFTSPPAVS SQSKLKQSN MLTAQEILHQ FGFTQTGIIS IERKPCNHQV
1051 DTAPTVALY DYTANRSDEL TIHRGDIIRV FFKDNEDWWY GSIGKGQGY
1101 FPAHVASET LYQELPPEIK ERSPPLSPEE KTKIEKSPAP QKQSINKNKS
1151 QDFRLGSESM THSEMRKEQS HEDQGHIMDT RMRKNKQAGR KVTLIE

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZpHtes3_in3, frame 1

TREMBL:U92792_1 gene: "Tup1"; product: "Tup1"; Schizosaccharomyces pombe general transcriptional repressor Tup1 (tup1) mRNA, complete cds., N = 1, Score = 186, P = 1e-10

TREMBL:AF104258_1 gene: "Pmc733"; product: "putative copper-inducible 35.6 kDa protein"; Festuca rubra putative copper-inducible 35.6 kDa protein (Pmc733) mRNA, complete cds., N = 1, Score = 235, P = 4.6e-18

TREMBL:SPAC3H5_8 gene: "SPAC3H5.08c"; product: "beta-transducin"; S.pombe chromosome I cosmid c3H5., N = 2, Score = 231, P = 2e-14

PIR:T02533 hypothetical protein F13M22.17 - Arabidopsis thaliana, N = 2, Score = 228, P = 1e-13

TREMBL:AF104258_1 gene: "Pmc733"; product: "putative copper-inducible 35.6 kDa protein"; Festuca rubra putative copper-inducible 35.6 kDa protein (Pmc733) mRNA, complete cds., N = 1, Score = 235, P = 4.6e-18

TREMBL:SPAC3H5_8 gene: "SPAC3H5.08c"; product: "beta-transducin"; S.pombe chromosome I cosmid c3H5., N = 2, Score = 231, P = 2e-14

TREMBL:CER03E1_1 gene: "R03E1.1"; Caenorhabditis elegans cosmid R03E1, N = 1, Score = 215, P = 2.3e-13

SWISSPROT:YZLL CAEEL HYPOTHETICAL 43.1 KD TRP-ASP REPEATS CONTAINING PROTEIN K04G11.4 IN CHROMOSOME X., N = 1, Score = 203, P = 7.1e-13

>TREMBL:AF104258_1 gene: "Pmc733"; product: "putative copper-inducible 35.6 kDa protein"; Festuca rubra putative copper-inducible 35.6 kDa protein (Pmc733) mRNA, complete cds.
Length = 321

HSPs:

Score = 235 (35.3 bits), Expect = 4.6e-18, P = 4.6e-18
Identities = 59/225 (26%), Positives = 111/225 (49%)

Query: 647 MRELCGHLNIIYDLSWSKDDHYILTSSSDGTARIWKNEINNTNTFRVLPHPSFVYTAKFH 706
+ E GH + I DLSWSK+ +L++S D T R+W ++ + +V H ++V +F+
Sbjct: 63 VHEFYGHGDAILDLSWSKNGD-LLSASMDKTVRLW--QVGRDSCLVKVESHTNYVTCVQFN 119

Query: 707 PAVRELVTGTCYDSMIRIWKVEMREDSAILVRQFDVHKSFINSLCFDTEGHHMYSGDCTG 766
P +TGC D ++RIW V LV + K + ++C+ +G +G TG
Sbjct: 120 PTNGNYFITGCDGLVRIWDVRK----CLVVDWANSKEIVTAVCYRPDGKAGVAGTITG 174

Query: 767 VIVVWNTYVKINDLEHSVHHWTINKKEIKETEFKGIPIISYLEIHPNGKRLLIHTKDSTLRI 826
++ +LE V ++N K + + Y P K+L++ + D+ +RI

Sbjct: 175 NCRYYDASENRLELESQV---SLNGRKSLHKRIVGFQYCPSPD--KKLMVTSGDAQVRI 229
 Query: 827 MDLRILVARKEVGAANYREKIHSTLTFCGTFLFAGSEGDGIVYVWN 871
 +D +++ + G + ++ + TP G + + S+D +Y+WN
 Sbjct: 230 LDGAHVISN-YKGLQS-SSQVARSFTPDGDHIVSASDDSRIMWN 272

Pedant information for DKFZphtes3_ln3, frame 1

Report for DKFZphtes3_ln3.1

[LENGTH] 1196
 [MW] 137114.70
 [pI] 6.79
 [HOMOL] SWISSPROT:YKY4_CAEEL HYPOTHETICAL 40.4 KD TRP-ASP REPEATS CONTAINING PROTEIN C14B1.4 IN CHROMOSOME III. 8e-21
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YKL121w] 2e-11
 [FUNCAT] 04.05.01.01 general transcription activities [S. cerevisiae, YBR198c TAF90 - TFIID subunit] 4e-10
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YBR198c TAF90 - TFIID subunit] 4e-10
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YPR178w] 1e-08
 [FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YPR178w] 1e-08
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YDR364c] 4e-08
 [FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YDR364c] 4e-08
 [FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL145c] 9e-08
 [FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae, YDL145c] 9e-08
 [FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YCR084c] 2e-07
 [FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YHL002w] 7e-07
 [FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YFR024c-a] 2e-06
 [FUNCAT] 02.16 fermentation [S. cerevisiae, YMR116c] 4e-06
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YMR116c] 4e-06
 [FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae, YMR116c] 4e-06
 [FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YFL009w] 4e-05
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YFL009w] 4e-05
 [FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YFL009w] 4e-05
 [FUNCAT] 03.01 cell growth [S. cerevisiae, YCR088w] 6e-05
 [FUNCAT] 03.25 cytokinesis [S. cerevisiae, YCR057c] 7e-05
 [BLOCKS] BL00024H
 [SCOP] dltbgd_ 2.46.3.1.1 betal-subunit of the signal-transducing 3e-91
 [SCOP] dlqfc_ 2.21.2.1.9 Growth factor receptor-bound protein 2 (GRB2), N 4e-14
 [SCOP] dlfmk_ 2.21.2.1.8 (1-64) c-src tyrosine kinase [human (Hom 5e-15
 [SCOP] dlad5b1_ 2.21.2.1.7 (1-63) Hemapoetic cell kinase Hck [human (Hom 3e-15
 [SCOP] dl1cka1_ 2.21.2.1.16 (1-54) p56-lck tyrosine kinase, SH3 domain [huma 1e-13
 [SCOP] dlqwea_ 2.21.2.1.15 Src kinase, SH3 domain [Avian sarcoma virus 2e-15
 [SCOP] dlshg_ 2.21.2.1.6 alpha-Spectrin, SH3 domain [chicken (Gallu 2e-13
 [SCOP] dlprmc_ 2.21.2.1.13 Src kinase, SH3 domain [chicken (Gallus gallus) 2e-15
 [SCOP] dlhsq_ 2.21.2.1.12 Phospholipase C, SH3 domain [human (Hom 2e-13
 [SCOP] dlabo_ 2.21.2.1.3 Abl tyrosine kinase, SH3 domain [Mouse (Mu 3e-13
 [SCOP] dlfn_ 2.21.2.1.2 Fyn, SH3 domain [human (Homo sapiens) 2e-15
 [SCOP] dlsema_ 2.21.2.1.11 Growth factor receptor-bound protein 2 (GRB2), N 1e-13
 [SCOP] dlgbqa_ 2.21.2.1.10 Growth factor receptor-bound protein 2 (GRB2), N 3e-16
 [SCOP] dlckaa_ 2.21.2.1.1 C-Crk, N-terminal SH3 domain [mouse (Mu 3e-15
 [EC] 3.1.4.3 Phospholipase C 2e-07
 [EC] 3.1.4.11 1-Phosphatidylinositol-4,5-bisphosphate phosphodiesterase 7e-07
 [EC] 3.6.1.32 Myosin ATPase 7e-07
 [EC] 2.7.1.112 Protein-tyrosine kinase 8e-06
 [PIRKW] nucleus 2e-08
 [PIRKW] phosphotransferase 8e-06
 [PIRKW] plasma 4e-07
 [PIRKW] duplication 4e-07
 [PIRKW] phosphoric diester hydrolase 2e-07
 [PIRKW] tandem repeat 7e-07
 [PIRKW] hormone 4e-07
 [PIRKW] transmembrane protein 2e-06
 [PIRKW] stomach 4e-07
 [PIRKW] actin binding 7e-07
 [PIRKW] ATP 7e-07
 [PIRKW] phosphoprotein 7e-07
 [PIRKW] signal transduction 7e-09
 [PIRKW] heterotrimer 7e-09
 [PIRKW] P-loop 7e-07
 [PIRKW] hydrolase 7e-07
 [PIRKW] transcription regulation 5e-06
 [PIRKW] GTP binding 7e-09

[SUPFAM] 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase II 2e-07
 [SUPFAM] SH3 homology 2e-07
 [SUPFAM] SH2 homology 2e-07
 [SUPFAM] protozoan myosin heavy chain IB 7e-07
 [SUPFAM] myosin motor domain homology 7e-07
 [SUPFAM] pleckstrin repeat homology 2e-07
 [SUPFAM] protein-tyrosine kinase src 8e-06
 [SUPFAM] WD repeat homology 3e-12
 [SUPFAM] 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase domain Y homology 2e-07
 [SUPFAM] protein kinase homology 8e-06
 [SUPFAM] 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase domain X homology 2e-07
 [SUPFAM] GTP-binding regulatory protein beta chain 7e-09
 [SUPFAM] yeast coatomer complex alpha chain 4e-07
 [PROSITE] RGD 1
 [PROSITE] MYRISTYL 6
 [PROSITE] AMIDATION 2
 [PROSITE] CAMP_PHOSPHO_SITE 4
 [PROSITE] CK2_PHOSPHO_SITE 25
 [PROSITE] TYR_PHOSPHO_SITE 4
 [PROSITE] PKC_PHOSPHO_SITE 19
 [PROSITE] ASN_GLYCOSYLATION 6
 [PFAM] Src homology domain 3
 [PFAM] WD domain, G-beta repeats
 [KW] Irregular
 [KW] 3D
 [KW] LOW_COMPLEXITY 5.77 %
 [KW] COILED_COIL 2.42 %

SEQ MPTAESEAKVKTKVRFKLLKTHSDLREKKKKKKLVRSEENISPDITIRSNLHYMKETT
 SEGxxxxxxx.....
 COILSCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....
 lgotB

SEQ SDDPDTIRSNLPHIKETTSDDVSAANTNNLKKSTRVTKNKLRLNTQLATENPNGDASVEED
 SEG
 COILS
 lgotB

SEQ KQGKPNKKVIKTVPQLTTQDLKPETPENKVDSTHQKTHTKPQPGVDHQKSEKANEGREET
 SEGxxx
 COILS
 lgotB

SEQ DLEEDEELMQAYQCHVTEEMAKEIKRKIRKKLKEQLTYFPSDTLFDHDKLSSEKRRKKKE
 SEG xxxxxxxx.....xxxxxxxxxxxxxxxx.....xxxxxxxxxxxx
 COILS
 lgotB

SEQ VPFVSKAETSTLTISGDTVEGEQKKESSVRSVSSDSHQDDEISSMEQSTEDSMQDDTKPK
 SEGxxxxxxxxxxxx.....xxx
 COILS
 lgotB

SEQ PKKTKKKTKAVADNNEDVDGVDGVEITSRDSVPYPKCLDDDLVLGVYIHRTRLKSDFM
 SEG xxxxxxxx.....
 COILS
 lgotB

SEQ ISHPMVKIHVVDEHTGQYVKKDDSGRPVSSYYEKENVDYILPIMTQPYDFKQLKSRLPEW
 SEG
 COILS
 lgotB

SEQ EEQIVFNENFPYLLRGSDSPKVLFFFEILDFLSVDEIKNNSEVQNQECGFRKIAWAFLK
 SEG
 COILS
 lgotB

SEQ LLGANGNANINSKRLQLYYPPTKPRSPLSVVEAFEWWSKCPRNHYPSTLYVTVRGLKVP
 SEG
 COILS
 lgotB

SEQ DCIKPSYRMMALQEEKGKPVHCERHHSSSSVDTEPGLEESKEVIKWRLPGQACRIPNK
 SEG
 COILS
 lgotB

```

SEQ      HLFSLNAGERGCFLDFSHNGRILAAACASRDGYPIILYEIPSGRFMRELCGHLNIIYDL
SEG      .....
COILS    .....
lgotB    .....CEEEEECECCCCCEEE

SEQ      SWSKDDHYILTSSSDGTARIWKNEINNTNTFRVLPHPSFVYTAKFHPAVRELVVTGCYDS
SEG      .....
COILS    .....
lgotB    EETTTTTEEEEEETTTEEEEEETT--TTCEEEEEETTCEEEEEETTT-TCEEEEEETTT

SEQ      MIRIWKVEMREDSAILVRQFDVHKFSFINSCLFDTEGHHMYSGDCTGVIVVWNTYVKINDL
SEG      .....
COILS    .....
lgotB    EEEEEETTTTBTTEEEEEEECCCCCE-EEEEEEETTEEEEEETTEEEEE.....

SEQ      EHSVHHWTINKEIKETEFKGIPIISYLEIHPNGKRLLIHTKDLRIMDLRILVARKFVGA
SEG      .....
COILS    .....
lgotB    .....

SEQ      ANYREKIHSTLTPCGTFLFAGSEGDIVYVWNPETGEQVAMYSDLFFKSPIRDISYHPFEN
SEG      .....
COILS    .....
lgotB    .....

SEQ      MVAFCAFQONEPILLYIYDFHVAQQAEMFKRYNGTFPLPGIHQSQDALCTCPKLPHQGS
SEG      .....
COILS    .....
lgotB    .....

SEQ      FQIDFVHTESSTKMQLVKQRLETVTEVIRSCAAKVNKNLSFTSPPAVSSQSKLKQSN
SEG      .....
COILS    .....
lgotB    .....

SEQ      MLTAQEILHQFGFTQTGIISIERKPCNHQVDTAPTVALDYDTANRDELTIHRGDIIRV
SEG      .....
COILS    .....
lgotB    .....

SEQ      FFKDNEDWWYGSIGKGQEGYFPANHVASETLYQELPPEIKERSPPLSPEEKTKIEKSPAP
SEG      .....
COILS    .....
lgotB    .....

SEQ      QKQSINKNKSQDFRLGSESMTHSEMRKEQSHEDQGHIMDTRMRKNKQAGRKVTLIE
SEG      .....
COILS    .....
lgotB    .....

```

Prosites for DKFZphtes3_in3.1

PS00001	460->464	ASN_GLYCOSYLATION	PDOC00001
PS00001	686->690	ASN_GLYCOSYLATION	PDOC00001
PS00001	934->938	ASN_GLYCOSYLATION	PDOC00001
PS00001	1000->1004	ASN_GLYCOSYLATION	PDOC00001
PS00001	1065->1069	ASN_GLYCOSYLATION	PDOC00001
PS00001	1148->1152	ASN_GLYCOSYLATION	PDOC00001
PS00004	91->95	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	264->268	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	305->309	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	1190->1194	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	48->51	PKC_PHOSPHO_SITE	PDOC00005
PS00005	66->69	PKC_PHOSPHO_SITE	PDOC00005
PS00005	93->96	PKC_PHOSPHO_SITE	PDOC00005
PS00005	170->173	PKC_PHOSPHO_SITE	PDOC00005
PS00005	232->235	PKC_PHOSPHO_SITE	PDOC00005
PS00005	268->271	PKC_PHOSPHO_SITE	PDOC00005
PS00005	304->307	PKC_PHOSPHO_SITE	PDOC00005
PS00005	327->330	PKC_PHOSPHO_SITE	PDOC00005
PS00005	352->355	PKC_PHOSPHO_SITE	PDOC00005
PS00005	384->387	PKC_PHOSPHO_SITE	PDOC00005
PS00005	440->443	PKC_PHOSPHO_SITE	PDOC00005
PS00005	533->536	PKC_PHOSPHO_SITE	PDOC00005
PS00005	546->549	PKC_PHOSPHO_SITE	PDOC00005
PS00005	643->646	PKC_PHOSPHO_SITE	PDOC00005
PS00005	677->680	PKC_PHOSPHO_SITE	PDOC00005
PS00005	690->693	PKC_PHOSPHO_SITE	PDOC00005
PS00005	702->705	PKC_PHOSPHO_SITE	PDOC00005

PS00005	823->826	PKC_PHOSPHO_SITE	PDOC00005
PS00005	973->976	PKC_PHOSPHO_SITE	PDOC00005
PS00006	22->26	CK2_PHOSPHO_SITE	PDOC00006
PS00006	59->63	CK2_PHOSPHO_SITE	PDOC00006
PS00006	77->81	CK2_PHOSPHO_SITE	PDOC00006
PS00006	116->120	CK2_PHOSPHO_SITE	PDOC00006
PS00006	137->141	CK2_PHOSPHO_SITE	PDOC00006
PS00006	180->184	CK2_PHOSPHO_SITE	PDOC00006
PS00006	245->249	CK2_PHOSPHO_SITE	PDOC00006
PS00006	276->280	CK2_PHOSPHO_SITE	PDOC00006
PS00006	283->287	CK2_PHOSPHO_SITE	PDOC00006
PS00006	288->292	CK2_PHOSPHO_SITE	PDOC00006
PS00006	292->296	CK2_PHOSPHO_SITE	PDOC00006
PS00006	327->331	CK2_PHOSPHO_SITE	PDOC00006
PS00006	390->394	CK2_PHOSPHO_SITE	PDOC00006
PS00006	454->458	CK2_PHOSPHO_SITE	PDOC00006
PS00006	510->514	CK2_PHOSPHO_SITE	PDOC00006
PS00006	570->574	CK2_PHOSPHO_SITE	PDOC00006
PS00006	663->667	CK2_PHOSPHO_SITE	PDOC00006
PS00006	672->676	CK2_PHOSPHO_SITE	PDOC00006
PS00006	804->808	CK2_PHOSPHO_SITE	PDOC00006
PS00006	985->989	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1023->1027	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1127->1131	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1132->1136	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1161->1165	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1170->1174	CK2_PHOSPHO_SITE	PDOC00006
PS00007	1083->1091	TYR_PHOSPHO_SITE	PDOC00007
PS00007	211->219	TYR_PHOSPHO_SITE	PDOC00007
PS00007	1083->1091	TYR_PHOSPHO_SITE	PDOC00007
PS00007	210->219	TYR_PHOSPHO_SITE	PDOC00007
PS00008	483->489	MYRISTYL	PDOC00008
PS00008	577->583	MYRISTYL	PDOC00008
PS00008	716->722	MYRISTYL	PDOC00008
PS00008	800->806	MYRISTYL	PDOC00008
PS00008	861->867	MYRISTYL	PDOC00008
PS00008	941->947	MYRISTYL	PDOC00008
PS00009	811->815	AMIDATION	PDOC00009
PS00009	1188->1192	AMIDATION	PDOC00009
PS00016	1074->1077	RGD	PDOC00016

Pfam for DKFZphtes3_ln3.1

HMM_NAME WD domain, G-beta repeats

HMM *MrGHnnWVWCVaFSPDGrWFIvSGSWDgTCRLWD*
 + GH+N +++++S D ++ I+++S DGT R+W

Query 650 LCGHLNIIYDLWSKDDHY-ILTSSSDGTARIWK 682

HMM_NAME Src homology domain 3

HMM *pyVIALYDYqAqdpDELSFkEGDIIiIEsDD.WWrgRnnnTNGQEGW
 P+V+ALYDY+A+++DEL++ +GDII + ++++ WW+G GQEG+
 Query 1054 PTVVALYDYTANRSDDELTIHRGDIIRVFFKDNDWWYGSIGK--GQEGY 1100

HMM IPSNYVEPi*
 +P+N V+ +

Query 1101 FPAHNVASE 1109

DKFZphtes3_20c21

group: testes derived

DKFZphtes3_20c21 encodes a novel 708 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

Sequenced by MediGenomix

Locus: /map="22q11.2-12.2"

Insert length: 3997 bp

Poly A stretch at pos. 3877, polyadenylation signal at pos. 3853

```

1  GGTAGGCGGG GCGGCGCGTG ACCTAAGGCC TCTCTGCCGC GCGCGCAGGT
51  ACGGGGCAGA AGTCGCAGGT ACCCAGCTGC TGCCACGTT TCTGGTCCAG
101 AGTCCCGAAC CCCGAGCACT GGGATGCCTG GCTACTCCGA GCCAAGGCAC
151 TGATGTTTGA ACTGGAACT TCAAAACGTT TAATAAGAGT CTTCAGGATG
201 GGTTTGAACT AGACAAGCTA GAAATTTCTT TAGAACACCA GCTCTAGCAT
251 GCATCTCCCA CTTTGGGCTT TCCTGGAGAG GAGCTTGAAG AGGTGGTTCT
301 GCAGACAGCC ACAGTGATAC TCAGGAAACC AGAGGAATGG ATTTGACTTT
351 TCTGCTAGGA TTCTTTGTTA TAGTTTCTCC CTGAGTTGTA AGAGGCATGG
401 AAATATACAT GAAACTGAAG AACCTGCAAG GAAGGGAAGT GGAACCTTCC
451 ATGCTGAGTG AAAACTAACC AAGTGGCAGT TGTGACTGAA AACACTGAAA
501 CCTACCACGT CCAGATTCAC TGGATTGGGG GATAGAGGAA CGGTCACAGC
551 TAGGGAGAAA GAAGTGATAC CGGAAAAGAA AACCTAAATG AAGAGAAATGA
601 GGATGACTGC ACAGTAGATG GCCACCTCTA CCTCCACAGA GGCAAAAGTCA
651 GCCTCGTGGT GGAATTATTT TTTCTTTAT GATGGTTCCA AGGTAAGGA
701 AGAAGGCGAT CCAACAAGAG CTGGCATTGG TTACTTTTAT CCTTCCCAGA
751 CCTGTGCTAG CCAACAGGAG TTGCTTTGTG GACAGATTGC TGGAGTTGTC
801 CGCTGTGTTT CTGACATTTC TGACTCTCCT CTTACTCTTG TTCGTCTGAG
851 AAAACTGAAG TTTGCCATAA AAGTTGATGG AGATTACCTT TGGGTGCTGG
901 GCTGTGCTGT GGAGCTCCCT GATGTCAGCT GCAAGCGGTT TCTGGATCAG
951 CTAGTTGGAT TCTTTAATTT TTACAATGGA CCTGTTTCCC TAGCTTATGA
1001 GAAGTGTCTC CAGGAAGAAC TGAGCACGGA GTGGGACACC TTCATCGAGC
1051 AAATTCGTAA AAACACCAGT GATCTGCATA AGATTTTCAA TTCCCTCTGG
1101 AACTTGGACC AAATAAAGT GGAGCCCTCG TTGTGCTGA AGGCAGCCCG
1151 CATTCTGCAG ACCTGCCAGC GCTCGCCTCA CATTCTCGCT GGCTGCATCC
1201 TCTATAAAGG ACTGATTGTC AGCACCCAAC TCCCGCCCTC CCTCACCGCC
1251 AAGGTCTCTG TTCAACGAAC AGCACCTCAG GAGCAGAGAC TCCCTACGGG
1301 AGGGGATGCC CCGCAGGAAC ATGGAGCGGC ATTGCCCCCG AATGTCCAGA
1351 TTTATCCCTG TTTTGTGACC AAAGAGGAAG CCATTAGTCT CCACGATTC
1401 CCGGTGGAAC AGATGACAAG GTCTCTAGCA TCTCCAGCAG GACTCCAGGA
1451 TGGTTAGGCC CAGCACCATC CAAAGGGTGG GAGCACATCT GCCCTGAAAG
1501 AAAACGCCAC TGGCCATGTG GAATCCATGG CCTGGACCAC CCAGATCC
1551 ACATCCCTCG ACGAAGCTTG TCCAGATGGC AGGAAGGAGA ACGGATGCTT
1601 GTCTGGCCAT GATCTGGAGA GCATCAGGCC CGCAGGACTG CACAACCTG
1651 CCAGGGGTGA GGTCTTGGC CTCAGCTCCT CCCTGGGGA GGAACCTAGT
1701 TTTCTCCAAG AAGAACTCGA CTTGTCTGAA ATCCACATTC CAGAGGCTCA
1751 GGAAGTGGA AATGGCTCAG GTCATTTTGC CTTCTACAT GTGCCTGTTT
1801 CAGATGGCAG GGCTCCTTAC TGCAAGGCAT CTCTCAGCGC CTCCAGCAGC
1851 CTGGAACCCA CGCTCCTGA GGACACAGCC ATCAGCAGCT TGGCCCTTCC
1901 CTCTGCTCCT GAGATGCTGA CCCAGCATGG AGCCCAAGAG CAGGTCGAAG
1951 ACCATCCTGG CCATAGCAGC CAAGCCCCCA TTCCCAGAGC AGACCTCTC
2001 CCCAGAAGGA CCCGACAGGC CTTGTTATTG CCTCGCTTAG ATCCAGGACA
2051 GAGAGGAAAC AAGCTTCCCA CGGGGGAACA AGGCCTGGAT GAGGATGTTG
2101 ATGGGGTCTG TGAAGGCCAC GCAGCCCTCT GTCTGGAATG CAGTTCAGGC
2151 TCAGCAAACG GTCAGGGTGC TGGCCCTCTC GCAGATGGAA TCAGCTCCAG
2201 CTGTACACCA GCAGAGTCTT GCATGGGGCT CGTGAGGATG AATCTCTACA
2251 CTCACCTGCGT CAAAGGGCTG ATGCTGTCCC TGCTGGCTGA GGAGCCGCTG
2301 CTGGGAGACA GCGCAGCCAT AGAGGAAGTG TACCACAGCA GCCTGGCTTC
2351 ACTGAATGGG CTGGAAGTCC ACCTGAAAGA GACGCTGCC AGGGATGAGG
2401 CAGCCTCCAC GAGCAGCACC TACAACCTCA CATATTACGA CCGCATTCAG
2451 AGCTTGCTGA TGGCAACCT GCCGCAGGTG GCCACCCCGC ATGATCGCCG
2501 CTTCTCTCAG GCGCTCAGCC TGATGCATAG CGAATTGGCC CAGCTGCCCG
2551 CGCTTTATGA AATGACTGTC AGAAATGCCT CCACGGCTGT GTACGCTGTG
2601 TGCAACCCCA TCCAGGAGAC ATATTTCCAG CAGCTGGCAC CTGCAGCAGC
2651 GAGCTCCGCG TTCCCAAACC CTCAGGATGG CGCCTTCAGC CTCTCCGGCA
2701 AAGCAAAGCA GAAGCTGCTG AAGCACGGGG TGAACCTGCT CTGAACTGCA
2751 CCCAGGAGGT GACTGGGAAG GAGAAAACCA GCAAAGGAAG CTCTGCCCTT
2801 TATAATTGAA AAGGCCCTC TATTTTATTT TTCTTGAAA CATTCCCTTT

```

```

2851 TTTAGGAACC AAATGATATT TGAGTTTTTG TTATTCCTTT TGCAGATTGG
2901 GATGTGTTTT GGGGGCAGGG GTTAGTTCTT CAGGTCGGCA GACCCAGAGC
2951 ACTTGATAAA GAACTGTATT TAATCGGTAG TGTTGGGGCC GGGACGGGCT
3001 TGCGTCCCTC TCTGCCATAC TGAGCCTGAG GTATTTTATA TCTCCTGCTG
3051 TTCCATCCCA GCTTGAATTG GTGCCACAAG CTTCGAAGTT GGCATTTTTT
3101 CTAGAACCCTG ATCGTCCACT AGCCGAGAGT GTGTGTGTTC AACCCCCACA
3151 CCAGGTGGTG GTAGGCGGTG TGACTGCACA GCGAGGTGCC GGATCTGTGA
3201 GCAGGCCGAC TCCACTCCCA CGCCGCAGGT AGGTTTCTCC AGTGCCTCT
3251 TGCTGGGAGG TCCGGATCGT TCCTGCAGGG AAGCGGCAGC ACACGGAGAC
3301 CACTTGGTTG AATTCTGTG GAACTCTACT CAAATCTAGG GCGTCTTCT
3351 TTGACCACAC AATGGGGGCA AGCCTTAATA ATATGGAAGG GAGTTTGGGC
3401 TTTAGAGATC CCTTTATAAA AGCTCTGGGG GCTGAGCCCT GAGAATTCAG
3451 TGACAACAGG ACCAACCTGC GCTGCCTTTG ACTACAAGTG GGCGTGCAG
3501 CTGGTTCCCTC TCGAGCGAGT GTCCCTAAAT AGGAGTTTAC AAGATGTCTG
3551 GGGGTAAAG CACTGTGCTT TTCAGTGGTG GCTGCGTGAA AGGGAGCGAC
3601 ACTCAGCTGT GTGTTCTTGG GCTTGTGTGG TACTTAGAAC CTCAGTTCTA
3651 TTACGTTATA GTCAGACATT TTTTGTACAG TATGAGACAG ACTGCAGGAT
3701 GAAATATTTT GTCAAAATCT TAACTGAATG TTTACTGGAA GTACTTGAGA
3751 TTCCATTGTA GAGTTGTATT GTTAATAATT TCATGTCAAGT GAACTGATAT
3801 CTGATGTTTA TGATATGGTG TCTTTTCTT GAAACAAGCT TCCAAGGGCT
3851 AGAATAAAAA TAGCCAAAAA ATGCTGGAAA AAAAAAAAAA AAAAAAAAAA
3901 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3951 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAA

```

BLAST Results

Entry HS1048E9 from database EMBLNEW:
 Human DNA sequence from clone 1048E9 on chromosome 22q11.2-12.2
 Contains pseudogene similar to ribosomal protein S3A and part of a gene
 similar to C.elegans protein CE02118, ESTs, STS, GSS.
 Score = 6540, P = 0.0e+00, identities = 1308/1308
 ~14 exons

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 618 bp to 2741 bp; peptide length: 708
 Category: putative protein
 Classification: no clue

```

1 MATSTSTEAK SASWNNYFFL YDGSKVKEEG DPTRAGICYF YPSQTLLDQQ
51 ELLCGQIAGV VRCVSDISDS PPTLVRLRKL KFAIKVDGDY LWVLGCAVEL
101 PDVSKRFLD QLVGFFNFYN GPVSLAYENC SQEELSTEDW TFIEQILKNT
151 SDLHKIENSL WNLDTQKVEP LLLKKAARIL QTCQRSFIL AGCILYKGLI
201 VSTQLPPSLT AKVLLHRTAP QEQLPTGGD APQEHGAALP PNVQIIPVTV
251 TKEEAI SLHE FPVEQMTSL ASPAGLQDGS AQHHPKGGST SALKENATGH
301 VESMAWTPD PTPDEACPD GRKENGCLSG HDLESIRPAG LHNSARGEVL
351 GLSSSLGKEL VFLQEELDLS EIHIPEAQEV EMASGHFAFL HVPVPDGRAP
401 YCKASLSASS SLEPTPPEDT AISSLRPSSA PEMLTQHGAQ EQVEDHPGHS
451 SQAPIPRADP LPRRTRRPLL LPRLDPGQRG NKLTGEQGL DEDVDGVCES
501 HAAPGLECSS GSANCOGAGP SADGISSRLT PAESCMGLVR MNLYTHCVKG
551 LMLSLLAEPP LLGDSAAIEE VYHSSLASLN GLEVHLKETL PRDEAASTSS
601 TYNFTYYDRI QSLLMANLPQ VATPHDRRFL QAVSLMHSEF AQLFALYEMT
651 VRNASTAVYA CCNPIQETYP QQLAPARSS GFNPQDGAF SLGKAKQKL
701 LKHGVNLL

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_20c21, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_20c21, frame 3

Report for DKFZphtes3_20c21.3

[illegible]

(No Prosite data available for DKFZphtes3 20c21.3)

(No Pfam data available for DKFZphtes3 20c21.3)

DKFZphtes3_20k2

group: signal transduction

DKFZphtes3_20k2 encodes a novel 839 amino acid protein with strong similarity to rat vanilloid receptor subtype 1.

VR1 seems to play an important role in the activation and sensitization of nociceptors. It is the receptor for e.g. capsaicin, a selective activator of nociceptors, a natural product of capsicum peppers. The novel protein is the human orthologue of rat VR1.

The new protein can find application as a target for the development of new nociception-modulating drugs.

strong similarity to rat vanilloid receptor subtype 1

Sequenced by MediGenomix

Locus: unknown

Insert length: 4187 bp

Poly A stretch at pos. 4154, polyadenylation signal at pos. 4135

```
1 GGCTCAGGCA GGCCTGGCCC AGAGTCACGC TGGCAACCAC GAGTTTGGGA
51 AGCAGTCGTA TTCTCTCTCT CTCTCTCTCT CTCTCAGTAT CCATGACAGT
101 GTGATGGAGA GTCTCTGCGG TGCCATCTGG GATGCAAACC GTCCTGTGT
151 CCCCCACGTC CAGGCCGTAG ATGCTCCCGG CCGGTCAGTC ACTTAGTCGT
201 CAGATCGCCC GTCTTGGTAT CACAGTGCTT CTGTTTAGGT TGCACACTGG
251 GCCACAGAGG ATCCAGCAAG GATGAAGAAA TGGAGCAGCA CAGACTTGGG
301 GGCAGCTGCG GACCACTCC AAAAGGACAC CTGCCAGAC CCCCTGGATG
351 GAGACCCTAA CTCCAGGCCA CCTCCAGCCA AGCCCCAGCT CTCCACGGCC
401 AAGAGCCGCA CCCGGCTCTT TGGGAAGGGT GACTCGGAGG AGGCTTTCCC
451 GGTGGATTGC CCTCAGAGG AAGGTGAGCT GGACTCCTGC CCGACCATCA
501 CAGTCAGCCC TGTATCACC ATCCAGAGGC CAGGAGACGG CCCCACCGGT
551 GCCAGGCTGC TGTCCAGGA CTCTGTCGCC GCCAGCACCG AGAAGACCCT
601 CAGGCTCTAT GATCGCAGGA GTATCTTTGA AGCCGTTGCT CAGAATAACT
651 GCCAGGATCT GGAGAGCCTG CTGCTCTTCC TGCAGAAGAG CAAGAAGCAC
701 CTCACAGACA ACGAGTTCAA AGACCCTGAG ACAGGGAAGA CCTGTCTGCT
751 GAAAGCCATG CTCACCTGCG ATGACGGACA GAACACCACC ATCCCCCTGC
801 TCCTGGAGAT CGCGCGGCAA ACGGACAGCC TGAAGGAGCT TGTCAACGCC
851 AGCTACACGG ACAGCTACTA CAAGGGCCAG ACAGCACTGC ACATCGCCAT
901 CGAGAGACGC AACATGGCCC TGGTGACCCT CCTGGTGGAG AACGGAGCAG
951 ACCTCCAGGC TGGCGCCCAT GGGGACTTCT TTAAGAAAAC CAAAGGGCGG
1001 CCTGGATTCT ACTTCGGTGA ACTGCCCTG TCCCTGGCCG CGTGCACCAA
1051 CCAGCTGGGC ATCGTGAAGT TCCTGCTGCA GAACTCCTGG CAGACGGCCG
1101 ACATCAGCGC CAGGGAATCG GTGGCAACA CGGTGCTGCA CGCCCTGGTG
1151 GAGGTGGCCG ACAACACGGC CGACAACACG AAGTTTGTGA CGAGCATGTA
1201 CAATGAGATT CTGATCCTGG GGGCCAAACT GCACCCGACG CTGAAGCTGG
1251 AGGAGCTCAC CAACAAGAAG GGAATGACGC CGCTGGCTCT GGCAGCTGGG
1301 ACCGGGAAGA TCGGGGTCTT GGCCTATATT CTCACGCGGG AGATCCAGGA
1351 GCCCGAGTGC AGGCACCTGT CCAGGAAGTT CACCGAGTGG GCCTACGGGC
1401 CCGTGCACTC CTCGCTGTAC GACCTGTCTT GCATCGACAC CTGCGAGAAG
1451 AACTCGGTGC TGGAGGTGAT CGCTACAGC AGCAGCGAGA CCCCTAATCG
1501 CCACGACATG CTCTTGGTGG AGCCGCTGAA CCGACTCCTG CAGGACAAGT
1551 GGGACAGATT CGTCAAGCGC ATCTTCTACT TCAACTTCCT GGTCTACTGC
1601 CTGTACATGA TCATCTTCAC CATGGCTGCC TACTACAGGC CCGTGGATGG
1651 CTTGCCCTCC TTTAAGATGG AAAAAATTGG AGACTATTTC CGAGTTACTG
1701 GAGAGATCCT GTCTGTGTTA GGAGGAGTCT ACTTCTTTT CCGAGGGATT
1751 CAGTATTTC TGCAGAGGCG GCCGTCGATG AAGACCCTGT TTGTGGACAG
1801 CTACAGTGAG ATGCTTTTCT TTCTGCAGTC ACTGTTATG CTGGCCACCG
1851 TGGTGTGTGA CTTAGCCAC CTCAAGGAGT ATGTGGCTTC CATGGTATTC
1901 TCCCTGGCCT TGGGCTGGAC CAACATGCTC TACTACACCC GCGGTTTCCA
1951 GCAGATGGGC ATCTATGCCG TCATGATAGA GAAGATGATC CTGAGAGACC
2001 TGTGCCGTTT CATGTTTGTC TACATCGTCT TCTTGTTTCG GTTTTCCACA
2051 GCGGTGGTGA CGCTGATTGA AGACGGGAAG AATGACTCCC TGCCGTCTGA
2101 GTCCACGTCG CACAGGTGGC GGGGGCCTGC CTGCAGGCCC CCCGATAGCT
2151 CCTACAAACG CCTGTACTCC ACCTGCCTGG AGCTGTTCAA GTTCAACATC
2201 GGCATGGGCG ACCTGGAGTT CACTGAGAAC TATGACTTCA AGGCTGTCTT
2251 CATCATCTCG CTGCTGGCCT ATGTAATTCT CACCTACATC CTCCTGCTCA
2301 ACATGCTCAT CGCCCTCATG GGTGAGACTG TCAACAAGAT CGCACAGGAG
2351 AGCAAGAACA TCTGGAAGCT GCAGAGAGCC ATCACCATCC TGGACACGGA
2401 GAAGAGCTTC CTTAAGTGCA TGAGGAAGGC CTTCCGCTCA GGCAAGCTGC
2451 TGCAAGTGGG GTACACACCT GATGGCAAGG ACGACTACCG GTGGTGCTTC
2501 AGGGTGGACG AGGTGAACCT GACCACCTGG AACACCAACG TGGGCATCAT
2551 CAACGAAGAC CCGGGCAACT GTGAGGGCGT CAAGCGCACC CTGAGCTTCT
2601 CCCTGCCGTC AAGCAGAGTT TCAGGCAGAC ACTGGAAGAA CTTTGGCCCTG
2651 GTCCCCCTTT TAAGAGAGGC AAGTGCTCGA GATAGGCAGT CTGCTCAGCC
```

```

2701 CGAGGAAGTT TATCTGCGAC AGTTTTCAGG GTCTCTGAAG CCAGAGGACG
2751 CTGAGGTCTT CAAGAGTCCT GCCGCTTCCG GGGAGAAAGT AGGACGTCAC
2801 GCAGACAGCA CTGTCAACAC TGGGCCTTAG GAGACCCCGT TGCCACGGGG
2851 GGCTGCTGAG GGAACACCAG TGCTCTGTCA GCAGCCTGGC CTGGTCTGTG
2901 CCTGCCAGC ATGTTCCCAA ATCTGTGCTG GACAAGCTGT GGAAGCGTT
2951 CTTGGAAGCA TGGGGAGTGA TGTACATCCA ACCGCTACTG TCCCAAGTG
3001 AATCTCCTAA CAGACTTCA GGTTTTACT CACTTTACTA AACAGTTTGG
3051 ATGGTCAGTC TCTACTGGGA CATGTTAGGC CCTTGTTC TTTGATTTTA
3101 TTCTTTTTTT TGAGACAGAA TTTCACTCTT CTCACCCAGG CTGGAATGCA
3151 GTGGCACAAT TTTGGCTCCC TGCAACCTCC GCCTCCTGGA TTCCAGCAAT
3201 TCTCTGCCT CGGCTTCCCA AGTAGCTGGG ATTACAGGCA CGTGCCACCA
3251 TGCTCTGGTA ATTTTTTGT TTTTTTTAAT AGATATGGGG TTTGCGCATG
3301 TTGGCCAGG TGGTCTCGAA CTCCTGACCT CAGGTGATCC GCCCACCTCG
3351 GCCTCCCAA GTGCTGGGAT TACAGGTGTG AGCCTCCACA CCTGGCTGTT
3401 TTCTTTGATT TTATTCTTTT TTTTTTTTCT GTGAGACAGA GTTCACTCT
3451 TGTGCCCAG GCTGGAGTGC AGTGGTGTGA TCTTGGCTCA CTGCAACCTC
3501 TGCTCCCGG GTTCAAGCGA TTCTTCTGCT TCAGTCTCCC AAGTAGCTTG
3551 GATTACAGGT GAGCACTACC ACGCCCGGCT AATTTTTGTA TTTTAATAG
3601 AGACGGGGT TCACCATGTT GGCCAGGCTG GTCTCGAATC CTTGACCTCA
3651 GGTGATCTGC CCGCCTTGGC CTCCCAAAGT GCTGGGATTA CAGGTGTGAG
3701 CCGCTGCGCT CGGCCTTCTT TGATTTTATA TTATTAGGAG CAAAAGTAAA
3751 TGAAGCCAG GAAAACACCT TTGGGAACAA ACTCTTCCTT TGATGGAAAA
3801 TGCAGAGGCC CTTCCTCTCT GTGCCGTGCT TGCTCCTCTT ACCTGCCCGG
3851 GTGTTTGGG GGTGTTGGTG TTTCTCCTT GGAGAAGATG GGGGAGGCTG
3901 TCCCACTCCC AGCTCTGGCA GAATCAAGCT GTTGACAGCAG TGCCTTCTTC
3951 ATCCTTCTCT ACGATCAATC ACAGTCTCCA GAAGATCAGC TCAATTGCTG
4001 TGCAGGTTAA AACTACAGAA CCACATCCCA AAGGTACCTG GTAAGAATGT
4051 TTGAAAGATC TTCCATTCTT AGGAACCCCA GTCCTGCTTC TCCGCAATGG
4101 ACATGCTTC CACTCCATCC ATACTGGCAT CCTCAAATAA ACAGATATGT
4151 ATACATATAA AAAAAAAAAA AAAAAAAAAA AAAAAA

```

BLAST Results

No BLAST result

Medline entries

99288727:

Recent advances in neuropharmacology of cutaneous nociceptors.

99231880:

A non-pungent triphenyl phenol of fungal origin, scutigerol, stimulates rat dorsal root ganglion neurons via interaction at vanilloid receptors.

Peptide information for frame 2

ORF from 272 bp to 2788 bp; peptide length: 839

Category: strong similarity to known protein

Classification: Cell signaling/communication

```

1 MKKWSSTDLG AAADPLQKDT CPDPLDGDPN SRPPPAKPQL STAKSRTLRF
51 GKGDSEEAFF VDCPHEEGEL DSCPTITVSP VITIQRPGDG PTGARLLSQD
101 SVAASTEKTL RLYDRRSIFE AVAQNNCQDL ESLLLFLQKS KKHLTDNEFK
151 DPETGKTCLL KAMLNLDHGO NTTIPLLEI ARQTDLSKEL VNASYTDSYY
201 KGQTAHIAI ERRNMAVTL LVENGADVQA AAHGDFFKKT KGRPGFYFGE
251 LPLSLAACTN QLGIVKFLLQ NSWQTADISA RDSVGNVTLH ALVEVADNTA
301 DNTKFVTSY NEILILGAKL HPTLKLELT NKKGMTPLAL AAGTGKIGVL
351 AYILQREIQE PECRHLRKF TEWAYGPVHS SLYDLSCIDT CEKNSVLEVI
401 AYSSSETPNR HDMLLVEPLN RLLQDKWDRF VKRIFYFNFL VYCLYMIIFT
451 MAAYYRPVDG LPPFKMEKIG DYFRVTGEIL SVLGGVYFFF RGIQYFLQRR
501 PSMKTLFVDS YSEMLFFLQS LFMLATVVLY FSHLKEYVAS MVFSLALGWT
551 NMLYYTRGFQ QMGIYAVMIE KMILRDLCRF MFVYIVFLFG FSTAVVTLIE
601 DGKNDLPSSE STSHRWGPA CRPPDSSYNS LYSTCLELFK FTIGMGDLEF
651 TENYDFKAVF IILLLAYVIL TYILLNMLI ALMGETVNKI AQESKNIWKL
701 QRAITILDE KSLFKCMRKA FRSGKLLQVG YTPDGKDDYR WCFRVDEVNW
751 TTWNTNVGII NEDPGNCEGV KRTLSFSLRS SRVSGRHWKN FALVPLLREA
801 SARDRQSAQP EEVYLRQFSG SLKPDAEVF KSPAASGEK

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFzphes3_20k2, frame 2

TREMBL:AF029310_1 product: "vanilloid receptor subtype 1"; Rattus norvegicus vanilloid receptor subtype 1 mRNA, complete cds., N = 1, Score = 3760, P = 0

TREMBLNEW:AB015231_1 product: "stretch-inhibitable nonselective channel (SIC)"; Rattus norvegicus mRNA for stretch-inhibitable nonselective channel (SIC), complete cds., N = 2, Score = 2090, P = 2e-219

>TREMBL:AF029310_1 product: "vanilloid receptor subtype 1"; Rattus norvegicus vanilloid receptor subtype 1 mRNA, complete cds.
Length = 838

HSPs:

Score = 3760 (564.1 bits), Expect = 0.0e+00, P = 0.0e+00
Identities = 721/839 (85%), Positives = 773/839 (92%)

```
Query:      1 MKKWSSTDGAAADPLQKDTCPDPLDGPNSRPPPAKPQLSTAKSRTLFGKGDSEEAFF 60
             M++ +S D + P Q+++C DP D DPN +PPP KP + T +SRTLFGKGDSEEA P
Sbjct:      1 MEQRASLDSEESPPQENSCLDPPDRDPNCKPPPVKPHIFTTRSRTLFGKGDSEEAFF 60

Query:     61 VDCPHEEGELDSCTITVSPVITIQRPDGPARGLLSQDSVAASTEKTLRLYDRRSIFE 120
             +DCP+EEG L SCP ITVS V+TIQRPDGP R SQDSV+A EK RLYDRRSIF+
Sbjct:     61 LDCPYEEGLASCPITVSSVLTIQRPDGPASVRPSSQDSVSAG-EKPPRLYDRRSIFD 119

Query:    121 AVAQNNQCQDLESLLFLQSKSKHLTDNEFKDPETGKTCLLKAMLNLDHGQNTTIPLLEI 180
             AVAQ+NCQ+LESLL FLQ+SKK LTD+EFKDPETGKTCLLKAMLNLDHGQNTTIPLLEI
Sbjct:    120 AVAQNNQCQDLESLLFLQSKSKHLTDNEFKDPETGKTCLLKAMLNLDHGQNTTIPLLEI 179

Query:    181 ARQTDLSKELVNASYTDSYKGTALHIAIERRNMALVTLLVENGADVQAAAHGDFFKKT 240
             AR+TDSLK+ VNASYTDSYKGTALHIAIERRNM LVTLLVENGADVQAAA+GDFFKKT
Sbjct:    180 ARQTDLSKELVNASYTDSYKGTALHIAIERRNMALVTLLVENGADVQAAAHGDFFKKT 239

Query:    241 KGRPGFYFGEPLSLAACTNQLGIVKFLQNSWQTADISARDSVGNTVLHALVEVADNTA 300
             KGRPGFYFGEPLSLAACTNQL IVKFLQNSWQ ADISARDSVGNTVLHALVEVADNT
Sbjct:    240 KGRPGFYFGEPLSLAACTNQLAIVKFLQNSWQADISARDSVGNTVLHALVEVADNTV 299

Query:    301 DNTKFVTSMYNEILILGAKLHPTLKLEELTNKKGMTPLALAAGTGKIGVLAYILQREIQE 360
             DNTKFVTSMYNEILILGAKLHPTLKLEE+TN+KG+TPLALAA +GKIGVLAYILQREI E
Sbjct:    300 DNTKFVTSMYNEILILGAKLHPTLKLEELTNKKGMTPLALAAGTGKIGVLAYILQREIHE 359

Query:    361 PECRHLSRKFTWAYGVPVHSSLYDLSCIDTCEKNSVLEVIAYSSSETPNRHDMMLLVEPLN 420
             PECRHLSRKFTWAYGVPVHSSLYDLSCIDTCEKNSVLEVIAYSSSETPNRHDMMLLVEPLN
Sbjct:    360 PECRHLSRKFTWAYGVPVHSSLYDLSCIDTCEKNSVLEVIAYSSSETPNRHDMMLLVEPLN 419

Query:    421 RLLQDKWDRFVKRIFYFNFLVYCLYMIIFTMAAYYRPVDGLPPFKMEK-IGDYFRVTGEI 479
             RLLQDKWDRFVKRIFYFNFLVYCLYMIIFT AAYYRPV+GLPP+K++ +GDYFRVTGEI
Sbjct:    420 RLLQDKWDRFVKRIFYFNFLVYCLYMIIFTMAAYYRPVEGLPPYKLNKTVGDYFRVTGEI 479

Query:    480 LSVLGGVYFFFRGIQYFLQRRPSMKTLFVDSYSEMLFFLQSLFMLATVVLYFSLKEYVA 539
             LSV GGVYFFFRGIQYFLQRRPS+K+LFVDSYSE+LFF+QSLFML +VVLYFS KEYVA
Sbjct:    480 LSVLGGVYFFFRGIQYFLQRRPSMKTLFVDSYSEILFFVQSLFMLVSVVLYFSQRKEYVA 539

Query:    540 SMVFSALGWTNMLYYTRGFQMGIIYAVMIEKMILRDLRCRFMFVYVFLFGFSTAVVTLI 599
             SMVFSLA+GWTNMLYYTRGFQMGIIYAVMIEKMILRDLRCRFMFVY+VFLFGFSTAVVTLI
Sbjct:    540 SMVFSALGWTNMLYYTRGFQMGIIYAVMIEKMILRDLRCRFMFVYVFLFGFSTAVVTLI 599

Query:    600 EDGKNDSLPESTSHRWGPACRPPDSSYNSLYSTCLELFKFTIGMGDLEFTENYDFKAV 659
             EDGKN+SLP EST H+ RG AC+P +SYNSLYSTCLELFKFTIGMGDLEFTENYDFKAV
Sbjct:    600 EDGKNDSLPESTSHRWGPACRPPDSSYNSLYSTCLELFKFTIGMGDLEFTENYDFKAV 658

Query:    660 FIILLAYVILTYILLNMLIALMGETVNKIAQESKNIWKLQRAITILDTEKSFLKCMRK 719
             FIILLAYVILTYILLNMLIALMGETVNKIAQESKNIWKLQRAITILDTEKSFLKCMRK
Sbjct:    659 FIILLAYVILTYILLNMLIALMGETVNKIAQESKNIWKLQRAITILDTEKSFLKCMRK 718

Query:    720 AFRSGKLLQVG+TPDGKDDYRWCFRVDEVNWTWNTNVIINEDPGNCEGVKRTLSFSLR 779
             AFRSGKLLQVG+TPDGKDDYRWCFRVDEVNWTWNTNVIINEDPGNCEGVKRTLSFSLR
Sbjct:    719 AFRSGKLLQVG+TPDGKDDYRWCFRVDEVNWTWNTNVIINEDPGNCEGVKRTLSFSLR 778

Query:    780 SSRVSGRHWKNFALVPLLRASARDRQSAQPEEVYLRQFSGSLKPDAEVFKSPAASGEK 839
             S RVSGR+WKNFALVPLLR+AS RDR + Q EEV L+ ++GSLKPDAEVFK GEK
Sbjct:    779 SSRVSGRHWKNFALVPLLRASARDRQSAQPEEVYLRQFSGSLKPDAEVFKSPAASGEK 838
```

Pedant information for DKFzphes3_20k2, frame 2

Report for DKF2phtes3_20k2.2

[LENGTH] 839
 [MW] 94950.75
 [pI] 6.90
 [HOMOL] TREMBL:AF029310_1 product: "vanilloid receptor subtype 1"; Rattus norvegicus
 vanilloid receptor subtype 1 mRNA, complete cds. 0.0
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YIL112w] 4e-05
 [PIRKW] alternative splicing 3e-06
 [PIRKW] peripheral membrane protein 3e-06
 [SUPFAM] ankyrin repeat homology 3e-06
 [SUPFAM] unassigned ankyrin repeat proteins 3e-06
 [PFAM] Ank repeat
 [KW] TRANSMEMBRANE 4

SEQ MKKWSSTDLGAAADPLQKDTCPDPLDGPNSRPPPAKQQLSTAKSRTLFGKGDSEEAFF
 PRD ccc
 MEM

SEQ VDCPHEEGELDSCTITVSPVITIQRPGDGPTGARLLSQDSVAASTEKTLRLYDRRSIFE
 PRD ccc
 MEM

SEQ AVAQNNQCQDLESLLLFLQSKKHLTDNEFKDPETGKTCLLKAMLNLDHGQNTTIPLLLEI
 PRD hhhhhcc
 MEM

SEQ ARQTDLSKELVNASYTDSYKQGTALHIAIERNMALVTLLVENGADVQAAAHGDFFKKT
 PRD hhhcc
 MEM

SEQ KGRPGFYFGEPLSLAACTNQLGIVKFLQNSWQTADISARDSVGNTVLHALVEVADNTA
 PRD ccc
 MEM

SEQ DNTKFVTSMYNEILILGAKLHPTLKLEELTNKKGMTPLAAGTGKIGVLAYILQREIQE
 PRD chhhc
 MEM

SEQ PECRHLRKFTEWAYGVPVHSSLYDLSCIDTCEKNSVLEVIAYSSSETPNRHOMLLVEPLN
 PRD ccc
 MEM

SEQ RLLQDKWDRFVKRIFYFNLVYCLYMIIFTMAAYRPPVDGLPPFKMEKIGDYFRVTGEIL
 PRD hhhc
 MEMMMMMMMMMMMMMMMMM.....

SEQ SVLGGVYFFFRGIQYFLQRRPSMKTFLFVDSYSEMLFFLQSLFMLATVVLYFSLKEYVAS
 PRD ccc
 MEMMMMMMMMMMMMMMMMM.....

SEQ MVFSLALGWTNMLYTRGFQOMGIYAVMIEKMILRDLCRFMFVYIVFLFGFSTAVVTLIE
 PRD hhhc
 MEMMMMMMMMMMMMMMMMM.....

SEQ DGKNDLSPSESTSHRWGPACRPPDSSYNSLYSTCLELKFETIGMGDLEFTENYDFKAVF
 PRD ccc
 MEM

SEQ IILLLAYVILTYILLNMLIALMGETVKNIAQESKNIWKLQRAITILDTEKSLKCMRKA
 PRD hhhc
 MEMMMMMMMMMMMMMMMMM.....

SEQ FRSGKLLQVGYTPDGKDDYRWCFRVDEVNWTNTNVTNGIINEDPGNCEGVKRTLSFSLRS
 PRD hhcc
 MEM

SEQ SRVSGRHWKNFALVPLLRASARDRQSAQPEEVYLRQFSGSLKPDAEVFKSPAASGEK
 PRD ccc
 MEM

(No Prosite data available for DKF2phtes3_20k2.2)

Pfam for DKF2phtes3_20k2.2

HMM_NAME	Ank repeat		
HMM	*GyTPLHIAARYNNvEMVr1LLQHGDIN*		
		G+T+LHIA +++N+ +V LL+++GAD+	
Query	202	GQTALHIAIERRNMALVTLLVENGADVQ	229

DKFZphtes3_2013

group: transmembrane protein

DKFZphtes3_2013 encodes a novel 595 amino acid protein with partial similarity to the IL-17 receptor.

The novel protein contains one transmembrane region.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes and as a new marker for testicular cells.

similarity to IL-17 receptor

Sequenced by MediGenomix

Locus: unknown

Insert length: 2406 bp

Poly A stretch at pos. 2345, no polyadenylation signal found

```
1 GCCTCAGGTG TTCCTGCGTT GTTGTGTCAGT GGAGAGCAGG GAGTGGGGCC
51 AGCCAGCAGA AACAGTGGGC TGTACAACAT CACCTTCAA TATGACAATT
101 GTACCACCTA CTTGAATCCA GTGGGGAAGC ATGTGATTGC TGACGCCAG
151 AATATCACCA TCAGCCAGTA TGCTTGCCAT GACCAAGTGG CAGTCACCAT
201 TCTTTGGTCC CCAGGGGCCC TCGGCATCGA ATTCTGAAA GGATTTCGGG
251 TAATACTGGA GGAGCTGAAG TCGGAGGGAA GACAGTGCCA ACAACTGATT
301 CTAAAGGATC CGAAGCAGCT CAACAGTAGC TTCAAAGAA CTGGAATGGA
351 ATCTCAACCT TTCCTGAATA TGAATTTGA AACGGATTAT TTCGTAAGG
401 TTGTCCTTTT TCCTTCCATT AAAACGAAA GCAATTACCA CCCTTTCTTC
451 TTTAGAACCC GAGCCTGTGA CCTGTTGTTA CAGCCGGACA ATCTAGCTTG
501 TAAACCTTTC TGAAGCCTC GGAACCTGAA CATCAGCCAG CATGGCTCGG
551 ACATGCAGGT GTCCTTCGAC CACGCACCGC ACAACTTCGG CTTCGGTTTC
601 TTCTATCTTC ACTACAAGCT CAAGCACGAA GGACCTTTCA AGCGAAAGAC
651 CTGTAAGCAG GAGCAAACTA CAGAGATGAC CAGCTGCCTC CTTCAAAATG
701 TTTCTCCAGG GGATTATATA ATTGAGCTGG TGGATGACAC TAACACAACA
751 AGAAAAGTGA TGCATTATGC CTTAAAGCCA GTGCACTCCC CGTGGGCCGG
801 GCCCATCAGA GCCGTGGCCA TCACAGTGCC ACTGGTAGTC ATATCGGCAT
851 TCGCGACGCT CTTCACTGTG ATGTGCCGCA AGAAGCAACA AGAAAATATA
901 TATTACATT TAGATGAAGA GAGCTCTGAG TCTTCCACAT AACTGCAGC
951 ACTCCCAAGA GAGAGGCTCC GGCCGCGGCC GAAGGTCTTT CTCTGCTATT
1001 CCAGTAAAGA TGGCCAGAAT CACATGAATG TCGTCCAGTG TTTCCGCTAC
1051 TTCCTCCAGG ACTTCTGTGG CTGTGAGGTG GCTCTGGACC TGTGGGAAGA
1101 CTTACGCTC TGTAGAGAAG GGCAGAGAGA ATGGGTATC CAGAAGATCC
1151 ACGAGTCCCA GTTCATCATT GTGGTTTGTT CCAAAGGTAT GAAGTACTTT
1201 GTGGACAAGA AGAACTACAA ACACAAAGGA GGTGGCCGAG GCTCGGGGAA
1251 AGGAGAGCTC TTCCTGGTGG CGGTGTCAGC CATTGCCGAA AAGCTCCGCC
1301 AGGCCAAGCA GAGTTCGTCC GCGGCGCTCA GCAAGTTTAT CGCCGTCTAC
1351 TTTGATTATT CCTGCGAGGG AGACGTCCCC GGTATCCTAG ACCTGAGTAC
1401 CAAGTACAGA CTCATGGACA ATCTTCTCA GCTCTGTTCC CACCTGCACT
1451 CCCGAGACCA CGGCCTCCAG GAGCCGGGGC AGCACACGCG ACAGGGCAGC
1501 AGAAGGAATC ACTTCCGGAG CAAGTCAGGC CGGTCCCTAT ACGTCGCCAT
1551 TTGCAACATG CACCAGTTTA TTGACGAGGA GCCCGACTGG TTCGAAAAGC
1601 AGTTCGTTCC CTTCCATCCT CCTCCACTGC GCTACCGGGA GCCAGTCTTG
1651 GAGAAATTTG ATTCCGGGCTT GGTTTTAAAT GATGTCATGT GCAAACCAGG
1701 GCCTGAGAGT GACTTCTGCC TAAAGGTAGA GCGGGCTGTT CTGGGGCAA
1751 CCGGACCAGC GACTCCAG CACGAGAGTC AGCATGGGGG CTTGGACCAA
1801 GACGGGGAGG CCCGGCCTGC CTTGACGGT AGCGCGGCC TGCAACCCCT
1851 GCTGCACACG GTGAAAGCCG GCAGCCCTC GGACATGCCG CGGGACTCAG
1901 GCATCTATGA CTCGTCTGTG CCCTCATCCG AGCTGTCTCT GCCACTGATG
1951 GAAGGACTCT CGACGGACCA GACAGAAACG TCTTCCCTGA CGGAGAGCGT
2001 GTCTCTCTCT TCAGGCCTGG GTGAGGAGGA ACCTCCTGCC CTTCTTCCA
2051 AGCTCCTCTC TTCTGGGTCA TGCAAAGCAG ATCTTGGTTG CCGCAGCTAC
2101 ACTGATGAAC TCCACGCGGT CGCCCTTTG TAACAAAACG AAAGAGTCTA
2151 AGCATTGCCA CTTTAGCTGC TGCCTCCCTC TGATTCCCA GCTCATCTCC
2201 CTGGTTGCAT GGCCCACTTG GAGCTGAGGT CTCATACAAG GATATTGGA
2251 GTGAATGCT GGCCAGTACT TGTCTCCCT TGCCCAACC CTTTACCGGA
2301 TATCTTGACA AACTCTCAA TTTTCTAAAA TGATATGGAG CTCTGAAAAA
2351 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2401 AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 346 bp to 2130 bp; peptide length: 595
 Category: similarity to known protein
 Classification: unclassified

```

1 MESQPFLLNMK FETDYFVKV VPFPSIKNESN YHPFFFRTRA CDLLQPDNL
51 ACKPFWKPRN LNISQHGSDM QVSFDHAPHN FGFRFFYLHY KKKHEGPFRK
101 KTCKQEQTTE MTSCLLQNV S PGDYIIEI LVD DTNTRKVMH YALKPVHSPW
151 AGPIRAVAIT VPLVVISAFA TLFTVMCRKK QQENIYSHLD EESSESSTYT
201 AALPRERLRP RPKVFLCYSS KDGQNHMNVV QCFAYFLQDF CGCEVALDLW
251 EDFSLCREGQ REWVIQKIHE SQFIIVVCSK GMKYFVDKKN YKHKGGRGS
301 GKGEFLVAV SAIAEKLRA QSSSAALSK FIAVYFDYSC EGDVPGILD
351 STKYRLMDNL POLCSHLHSR DHGLQEPGQH TRQGSRRNYF RSKSGRSLYV
401 AICNMHQFID EEPDWFEKQF VPFHPPPLRY REPVLEKFD S GLVLNDVMCK
451 PGPESDFCLK VEAALVGATG PADSQHESQH GGLDQDGEAR PALDGSAAALQ
501 PLLHTVKAGS PSDMPRDSGI YDSSVPSSSEL SLPLMEGLST DQTETSSLTE
551 SVSSSSGLGE EEPALPSKL LSSGSKADL GCRSYTDELH AVAPL

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2013, frame 1

TREMBL:U58917_1 product: "IL-17 receptor"; Homo sapiens IL-17 receptor
 mRNA, complete cds., N = 1, Score = 215, P = 4.7e-14

TREMBL:MM31993_1 product: "interleukin 17 receptor"; Mus musculus
 interleukin 17 receptor mRNA, complete cds., N = 2, Score = 152, P =
 1.1e-13

>TREMBL:U58917_1 product: "IL-17 receptor"; Homo sapiens IL-17 receptor
 mRNA, complete cds.
 Length = 866

HSPs:

Score = 215 (32.3 bits), Expect = 4.7e-14, P = 4.7e-14
 Identities = 85/284 (29%), Positives = 131/284 (46%)

```

Query: 213 KVFLCYSSKDGQNHMNVVQCFAYFLQDFCGCEVALDLWEDFSLCREGQREWV-IQK---I 268
      KV++ YS+ D +++VV FA FL CG EVALDL E+ ++ G WV QK +
Sbjct: 379 KVMIIYSA-DHPLYVDVVLKFAQFLLTACGTEVALDLLEEQAISEAGVMTWVGRQKQEMV 437

Query: 269 HESQFIIVVCSKGMKY---FVDKKNXXXXXXXXXXXXXELFLVAVSATAEXXXXXXXXXX 324
      + IIV+CS+G + + + +LF A++ I
Sbjct: 438 ENSKIIIVLCSRGRTRAKWQALLGRGAPVRLRCDHGKPVGDLFTAAMNMILPDFKRPACFG 497

Query: 325 XXXXXXFIAYVF-DYSCGDVPGILDSTKYRLMDNLPOLCSHLHSRDHGLQEPGQHTRQ 383
      ++ YF + SC+GDVP + + +Y LMD ++ + +D + +PG+ R
Sbjct: 498 T-----YVVCYFSEVSCDGDVPLFGAAPRYPLMDRFEEV--YFRIQDLEMFPQGRMHRV 550

Query: 384 G--SRRNYFRSKSGRSLYAICNMHQFIDEEDWFEKQFV----PFHPPPLR---YREPV 434
      G S NY RS GR L A+ + PDWFE + + P L + EP+
Sbjct: 551 GELSGDNYLRSPGGRLRAALDRFRDQVRCPDWFECENLYSADDQDAPSLDEEVFEEPL 610

Query: 435 LEKFDSDLVLNDVMCKPGPESDFCLKVEAAVLGATGPADSQHESQHGGLDQDGEAR 491
      L +G+V + + P S CL ++ V G G A H L G+ P
Sbjct: 611 LPP-GTGIVKRAPLVRE-PGSQACLAIDPLV-GEEGGAATAKLEPH--LQPRGQPAP 662

```

Pedant information for DKFZphtes3_2013, frame 1

Report for DKFZphtes3_2013.1

[LENGTH] 595
 [MW] 66847.05
 [pI] 6.27
 [HOMOL] TREMBL:MM31993_1 product: "interleukin 17 receptor"; Mus musculus interleukin
 17 receptor mRNA, complete cds. 2e-14
 [BLOCKS] BL00740A MAM domain proteins
 [BLOCKS] BL01224B N-acetyl-gamma-glutamyl-phosphate reductase proteins
 [KW] TRANSMEMBRANE 1
 [KW] LOW_COMPLEXITY 13.61 %

SEQ MESQPFLNMKFETDYFVKVVPFPSIKNESNYHPFFFRTRACDLLLPDNLACKPFWKPRN
 SEG
 PRD ccc
 MEM

SEQ LNISQHGSDMQVSFDHAPHNFGFRFFYLHYKLKHEGPFKRKTCKQEQTTEMTSCLLQNVN
 SEG
 PRD eeeeecc
 MEM

SEQ PGDYIIELVDDTNTTRKVMHYALKPVHSPWAGPIRAVAITVPLVVISAFATLFTVMCRKK
 SEG
 PRD ccc
 MEMMMMMMMMMMMMMMMMM.....

SEQ QQENIYSHLDEESSESSTYTAALPRERLRPRPKVFLCYSSKDGQNHMNVVQCFAFLQDF
 SEG
 PRD hhhhhhhhhcc
 MEM

SEQ CGCEVALDLWEDFSLCREGQREWVIQKIHESQFIIVVCSKGMKYFVDKKNYKHKGGRGS
 SEG
 PRD cchhhhhhhhhcc
 MEM

SEQ GKGELFLVAVSAIAEKLQAKQSSSAALSKEFIYVFDYSCEGDVPGILDSTKYRLMDNL
 SEG
 PRD ccc
 MEM

SEQ PQLCSHLHRDHLQEPGQHTROGSRNRYFRSKSGRSLYVAICNMHQFIDEEDWFEKQF
 SEG
 PRD cchhhhhhhcc
 MEM

SEQ VPFHPPLRYREPVLKFDGLVLDVMCKPGPESDFCLKVEAAVLGATGPADSQHESQH
 SEG
 PRD ecc
 MEM

SEQ GGLDQDGEARPALDGSAAQLPLHTVKAGSPSDMPRDSGIYDSSVPSSELSLPLMEGLST
 SEG
 PRD ccc
 MEM

SEQ DQTETSSLTESVSSSSGLGEEPPALPSKLLSSGSKADLGCRSYTDELHAVAPL
 SEG
 PRD hhhhhhhhhhecc
 MEM

(No Prosite data available for DKFZphtes3_2013.1)

(No Pfam data available for DKFZphtes3_2013.1)

DKFZphtes3_20ml8

group: nucleic acid management

DKFZphtes3_20ml8 encodes a novel 132 amino acid protein with similarity to the *S. cerevisiae* mitochondrial carrier protein RIM2.

The novel protein contains a leucine zipper and a Prosite mitochondrial energy transfer proteins signature. It is member of a family of substrate carrier proteins which are found in the inner mitochondrial membrane and are involved in energy transfer. The RIM2/MRS12 gene encodes a predicted protein of 377 amino acids that is essential for mitochondrial DNA metabolism and proper cell growth. Inactivation of this gene causes the total loss of mitochondrial DNA and, compared to wild-type rho⁰ controls, a slow-growth phenotype on media containing glucose. The novel protein seems to be the human orthologue of this protein.

The new protein can find application in modulation of mitochondrial DNA replication and maintenance.

similarity to carrier protein RIM2

Sequenced by MediGenomix

Locus: unknown

Insert length: 3572 bp

Poly A stretch at pos. 3530, polyadenylation signal at pos. 3510

```

1  GCCGCGGGGA GGGCTGTGCC GGTGCTTTC TGCAGCCGCA TCTCGGCCAG
51  CTCTCCTCGC CGTCCCCGGG GCGCTGTGCG TCTCCAGTCC GGGACCGAAG
101 CCGCCTGCCG TAGCGGGCGG CCAGATCCGC GTCCCGCCTC AGCGGCCGGA
151 GGACATGCCG GAGAGAGAAT GAGCCAGAGG GACACGCTGG TGCATCTGTT
201 TGCCGGAGGA TGTGGTGGTA CAGTGGGAGC TATTCTGACA TGTCCTACTGG
251 AAGTTGTAAA AACACGACTG CAGTCATCTT CTGTGACGCT TTATATTCTT
301 GAAGTTCAGC TGAACACCAT GGCTGGAGCC AGTGTCAACC GAGTAGTGTC
351 TCCCGGACCT CTTTATTGCC TAAAGGTGAT CTTGGAAAAA GAAGGGCCTC
401 GTTCCTTGTT TAGAGGACTA GGCCCCAATT TAGTGGGGGT AGCCCCCTCC
451 AGAGCAATAT ACTTTGCTGC TTATTCAAAC TGCAAGGAAA AGTTGAATGA
501 TGTATTTGAT CCTGATTCTA CCCAAGTACA TATGATTTCA GCTGCAATGG
551 CAGGTATGAA TGTATAATAT TAAAAAATAA AAAAAGTTTC TGAAACCTAG
601 AGGCTTAATA TTGAATTATA AGTTTGTAGT GAAAAGTTGA TGATTAAATGT
651 GCTTTTCATT GATTAGATGA TTTTACGTT TATCGATATA AACCAAATTA
701 GGTATATGTA AAATCTGTCA TCAGTTGACA TTTTGTAGT CAGGAGTTTA
751 CATGCTAGGG TACAAGTAAT ATATTATAT TGCCTTGTGT AGTCCACTGA
801 ATGTTTAGTG ATCATTGTGA ACAGTTTAA GAATCCAACC ATAATTACAC
851 TATAAATAAG TTATGGAGCT GTAATTTACT CTCTCTCCTT CAATTTCTGT
901 TAGTGCCCTT TCCCTTTTGT CTGCATGTTT TGGCTTCTGT CTGAAATGTG
951 TCGGCAATTC TTGGTAAAGT ATTCATTTTG TCCTGTGCTC AAATGCTGAA
1001 ATTTTGTGA GTGATGTATT ATTATTGACA ATTCAGTTAC TATGTGTATT
1051 TTTTAAAAAT GTTTATTATT CTACATAATT CACACTAGAC AGCACCTGAA
1101 ATTTAGACAC TGGCTATGTG TACATGCTTA CTATAGAAAT GTTTCAGGAA
1151 ACTCTCTGTT TCTGTCATCA CTGATAAGTA TATATGATT TGAATTAATA
1201 TAACTAGTTT TAGGTCTTTA CCCTGCCATA AAGATAAACA GTTGGTTTGA
1251 CCAATCTGCT TCTGGAATCA TTTGCTGCTA TGCATGTTAG ACAAGGCCAC
1301 GAACCTTGAT TTTCCATTGA AAATTCTCCC TAATATCTGA GATTATTATT
1351 ATATTACTAC ATATCTCACA TTTTCAAATT ATGCTGTAAC TTTATAAACT
1401 GTAGCTGCTT TCATCAGCTA TTGATCAATA AATTGAATGT CAATTATGTG
1451 CTTAATAATG AGTGCCTTAA ACTGTAAAC ACTTTTGGTT TAGAATAATA
1501 GTGAATCAAT TTGACCTATA TACTTCATGA AGTAAGTAAG TTGAAATAC
1551 AAATTTCTGA AAGGTCAATA GCCCTTATCG TATTACAAT TGTTTTAAAG
1601 GCTTTTGTGA TTTATTAATT GTCAGTTGAT TCACTGAAGC TTTAAACTG
1651 GAAGGGACAA TCCAAAGGTC AAAAGAGTGA AATACAATCA TTTACCAATA
1701 AGGAAACCTT GGGCAAATTA TGTAAATTTAT GTGAACCTCT CTTAGCTTAC
1751 CCATGGAATG AGTCAAGTGG TCTACATAGA TTTGGATTTT GAGAATTAGT
1801 TCTTTCATTT AGTGTATAG AGATTATCTT GTTACAACCTA GAATTATTTT
1851 TAATGTAATT TTTACAGATG TTGAATATTA GTAGATAGGA TTTTCCCTCT
1901 ACGAATTTGG ATGTAAGGTA AAGGTTGGTG GCCAGTGACA AACCTTATAA
1951 CCACCTTTAT AGGTCTTTTA AAAATATATT TGTGAATTAC CAGTGATTAT
2001 GTTTTGGCT TATAACCTCA GATAATTATA AAGAAATGTT AATCTTATTT
2051 GAAAGAATTG GAATCTAGAA AGTTAGATGA GCAGTCATTT TATATTGATA
2101 TTTGTATAT CAGTATAGCA AATGCAGAGG TTCAGAAAT CTTTATTTCC
2151 ACTGGAACAT CTTATTTTCA TAGAGTATCT CATCAGAATT TATTACTGTA
2201 TTTGTATCAC ATTGCAAAGA ATTCAGTAG AATTGTCAGT TTGCACTTTT
2251 TTCTCAAATG TGTACAAATG TTAACATATA GTTCATTTT ATCTGTACAT
2301 TGATGCCATT TCCCAACTTG AATTCCTCAA GTTTTGGTAA ACTTACAATC
2351 TCATACTTGT TCAGAGGTTA TTGCACTGTA CACTTACTGT GTAGAAAATA
2401 CTGTTTGAAT TTGTTTGCAG TTACATTGTT CTGAGAACTG TGCTCTCAGA
2451 GCTTCTGTCG ACTATTCATG AGCATTAAAC CTTAGCCTTG CAGTTTATTA

```

```

2501 CATAACTATA TGGTTAGTAA AACTGAATGG TCCAATGCAG ACTCATTAAA
2551 GTAGGCTTTT GCCCCTTTG TTCTTGAAAT AATCTAGACC AGATTACTCG
2601 GGGTTTTTTT TAGGATTATT TTTATAGGTC TAAATATGAA TGATTGGGG
2651 GTATGAAGTA CTAAAGATA GTTCTGTGAA AAATCATTIT CAGCTGTCTA
2701 TTCAAGGGAA AAAATGCTAA CCTTGTCAC TACTACACA AAACCACACT
2751 AAAATAAACCC ATTAATGATA CTGCCTGCAA GATTTTAAAC CACCAGATAG
2801 CACACACATT AAGGATTTAT AAGGCACTGT ACGTAATTTT TATCCAAGT
2851 GACCTCTCAA TTCATTTTCA TTTTGCAATT TATCCATATG AACTCATGTT
2901 TAATTAGAT AATAAAAAAT TATTTATTA AAAGGACAGT TTATTAAAG
2951 TGGGTCTTTT TATTTGTTGT AGTGCATACT ATAAGAATTI GTAAGCCTCT
3001 AAAGTTGAGC TATAAATTTT CATGCATTAA AAATTTGTTT CAGTTGTGAG
3051 GATATTTAAT CAGATTAAAT AATGTTGACT CTTAATATTT TGCCTGCCTT
3101 TTTTCTCTCC TACACATGAC CTTTGACAGA CTAAGTATAT CTCAGCTATT
3151 GAGGGTATCT GTTTTGTGTC CTGTATATTT TGTTTAAATT AACTTGTATA
3201 TTCCTTTGTA TACACCTAGG CACAGATGTA TGCAAAAAAA ATTTGTTAAA
3251 TTAATCTCTT CTTTATACTA ATCTCAATT TTTAAAGAT TTTATCTGGC
3301 ATGTATATAC TTTTATATAG AACATTATAA ATGTAAAGGA AATGAATCT
3351 AATTTTAATT GGATTATGTA TTCATACAGT TATCTCAAT TTTTAAATA
3401 CTAATAATGT AATCATTGAA TGTTTCCTAC ATACGTAGTG GGTTTTATTT
3451 GCTCACAGCA TACAGTTATT TTCAATTTA TGTTTTCTA TTAGACTTAA
3501 ATTTTCATTAT AATAAAGGCT TTTACTCATT AAATACAAAA AAAAAAAAAA
3551 AAAAAAAAAA AAAAAAAAAA AA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

95198680:

Overexpression of a novel member of the mitochondrial carrier family rescues defects in both DNA and RNA metabolism in yeast mitochondria.

Peptide information for frame 1

ORF from 169 bp to 564 bp; peptide length: 132
 Category: similarity to known protein
 Classification: Intracellular transport and traffic
 Prosite motifs: LEUCINE_ZIPPER (27-49)
 MITOCH_CARRIER (26-36)

```

1 MSQRDTLVHL FAGGCGGTVG AILTCPLEVV KTRLQSSSVT LYISEVOLNT
51 MAGASVNRVV SPGPLHCLKV ILEKEGPRSL FRGLGPNLVG VAPSRAIYFA
101 AYSNCKEKLN DVFDPDSTQV HMISAAMAGM NV

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_20m18, frame 1

PIR:S44092 probable carrier protein c2 - *Caenorhabditis elegans*, N = 2,
 Score = 147, P = 1.5e-19

PIR:S36081 probable carrier protein RIM2, mitochondrial - yeast
 (*Saccharomyces cerevisiae*), N = 1, Score = 230, P = 6.2e-19

>PIR:S36081 probable carrier protein RIM2, mitochondrial - yeast
 (*Saccharomyces cerevisiae*)
 Length = 377

HSPs:

Score = 230 (34.5 bits), Expect = 6.2e-19, P = 6.2e-19
 Identities = 55/133 (41%), Positives = 80/133 (60%)

```

Query:      8 VHLFAGGCGGTVGAILTCPLEVVKTRLQSSS-VTLYISEVOLNTMAGA----SVNRVVSP 62
              VH AGG GG GA++TCP ++VKTRLQS  + Y S+ +N  G+  S+N V+
Sbjct:     54 VHFVAGGIGGMAGAVVTCFDFLVKTRLQSDIFLKAYKSQA-VNISKGSTRPKSINYVIQA 112

```

Query: 63 GP-----LHCLKVILEKEGPRSLFRGLGPNLVGVAPSRAIYFAAYSNCCKELNDVFD--P 115
 G L + + ++EG RSLF+GLGPNLVGV P+R+I F Y K+ F+
 Sbjct: 113 GTHFKETLGIIGNVYKQEGFRSLFKGLGPNLVGVIPARSINFFTYGTTKDMYAKAFNNGQ 172

Query: 116 DSTQVHMISAAMAG 129
 ++ +H+++AA AG
 Sbjct: 173 ETPMIHLMAAATAG 186

Score = 77 (11.6 bits), Expect = 1.1e+00, P = 6.8e-01
 Identities = 25/88 (28%), Positives = 39/88 (44%)

Query: 3 QRDITLVHLFAGGCGGTVGAILTCPLEVVKTRLQSSSVTLYISEVQLNTMAGASVNRVVS 62
 Q ++HL A G A T P+ ++KTR VQL+ SV + +
 Sbjct: 172 QETPMIHLMAAATAGWATATATNPIWLIKTR-----VQLDKAGKTSVRQYKNS 219

Query: 63 GPLHCLKVILEKEGPRSLFRGLGPNLVG 90
 CLK ++ EG L++GL + +G
 Sbjct: 220 WD--CLKSVIRNEGFTGLYKGLSASYLG 245

Score = 71 (10.7 bits), Expect = 6.6e+00, P = 1.0e+00
 Identities = 28/91 (30%), Positives = 45/91 (49%)

Query: 12 AGGCGGTVGAILTCPLEVVKTRLQSSSVTLYISEVQLNTMAGASVNRVVS 71
 + G V +I T P EVV+TRL+ + + N G R + G + KVI
 Sbjct: 294 SAGLAKFVASIATYPHEVVRTLRQTP-----KEN---G---KRKYT-GLVQSFKVI 338

Query: 72 LEKEGPRSLFRGLGPNLVGVAPSRAIYFAAY 102
 +++EG S++ GL P+L+ P+ I F +
 Sbjct: 339 IKEGLFSMYSGLTPLHMRTPVNSIIMFGTW 369

Pedant information for DKFZphtes3_20m18, frame 1

Report for DKFZphtes3_20m18.1

[LENGTH] 132
 [MW] 13993.36
 [pI] 8.42
 [HOMOL] PIR:S36081 probable carrier protein RIM2, mitochondrial - yeast (Saccharomyces cerevisiae) 7e-19
 [FUNCAT] 07.16 purine and pyrimidine transporters [S. cerevisiae, YBR192w] 3e-20
 [FUNCAT] 08.04 mitochondrial transport [S. cerevisiae, YBR192w] 3e-20
 [FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YBR192w] 3e-20
 [FUNCAT] 02.13 respiration [S. cerevisiae, YBR192w] 3e-20
 [FUNCAT] 01.05.07 carbohydrate transport [S. cerevisiae, YPR021c] 3e-10
 [FUNCAT] 07.07 sugar and carbohydrate transporters [S. cerevisiae, YPR021c] 3e-10
 [FUNCAT] 07.99 other transport facilitators [S. cerevisiae, YEL006w] 1e-09
 [FUNCAT] 01.07.10 transport of vitamins, cofactors, and prosthetic groups [S. cerevisiae, YIL006w] 3e-09
 [FUNCAT] 07.04.07 anion transporters (cl, so4, po4, etc.) [S. cerevisiae, YKL120w] 2e-08
 [FUNCAT] 01.03.19 nucleotide transport [S. cerevisiae, YPR011c] 3e-08
 [FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YKR052c] 4e-08
 [FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YJR095w] 2e-07
 [FUNCAT] 01.01.07 amino-acid transport [S. cerevisiae, YOR130c] 5e-05
 [FUNCAT] 07.10 amino-acid transporters [S. cerevisiae, YOR130c] 5e-05
 [FUNCAT] 01.04.07 phosphate transport [S. cerevisiae, YJR077c] 7e-05
 [FUNCAT] 13.04 homeostasis of other ions [S. cerevisiae, YJR077c] 7e-05
 [BLOCKS] BL00215B Mitochondrial energy transfer proteins
 [BLOCKS] BL00215A Mitochondrial energy transfer proteins
 [PIRKW] duplication 6e-09
 [PIRKW] transmembrane protein 6e-09
 [PIRKW] mitochondrial inner membrane 4e-07
 [PIRKW] transport protein 5e-06
 [PIRKW] mitochondrion 7e-08
 [PIRKW] chloroplast 3e-08
 [SUPFAM] Btl protein 3e-08
 [SUPFAM] ADP,ATP carrier protein repeat homology 4e-09
 [SUPFAM] Caenorhabditis probable carrier protein c2 4e-09
 [SUPFAM] probable carrier protein YPR021c 6e-09
 [PROSITE] LEUCINE_ZIPPER 1
 [PROSITE] MITOCH_CARRIER 1
 [PFAM] Mitochondrial carrier proteins
 [KW] Alpha_Beta

SEQ MSQRDTLVHLFAGGCGGTVGAILTCPLEVVKTRLQSSSVTLYISEVQLNTMAGASVNRVV

```

PRD      cccccceeeccccccccceeeecchhhhhhhhhhhcccccccccccccccccccccc
SEQ      SPGPLHCLKVILEKEGPRSLFRGLGPNLVGVAPSRAIYFAAYSNCKEKLNDVFDPDSTQV
PRD      cccchhhhhhhhhhhccccceeeccccceeeccccceeeehhhhhhhhhcccccccccc
SEQ      HMISAAMAGMNV
PRD      chhhhhhhcccc

```

Prosite for DKFZphtes3_20ml8.1

```

PS00029      27->49      LEUCINE_ZIPPER      PDOC00029
PS00215      26->36      MITOCH_CARRIER      PDOC00189

```

Pfam for DKFZphtes3_20ml8.1

```

HMM_NAME      Mitochondrial carrier proteins
HMM            *pFwkDFLAGGIAGmMeHTvMFPIDtIKTRMQlQgEMpM..ahpR.....
            ++++++AGG +G + +++++P++++KTR+Q++ ++ + ++
Query          5  DTLVHLFAGGCGGTVGAILTCPLEVVKTRLQSS-SVTLYISEVQLNTMA      52
HMM            .....YkGMIdCFRwiwkNEGWRGLWRGLgANvIRYIPqWaIRFGFY
            G+++C++ I+++EG+R+L+RGLG+N++++P +AI+F+ Y
Query          53 GASVNRVSPGPLHCLKVILEKEGPRSLFRGLGPNLVGVAPSRAIYFAAY      102
HMM            EFMKeMFiDyfgeddnYwMwFwmnYMaGs*
            +KE ++D F++ D+++++ + +MAG+
Query          103 SNCKEKLNDVFDp-DSTQVHMISAAMAGM      130

```


DKFZphtes3_21d4

group: signal transduction

DKFZphtes3_21d4 encodes a novel 464 amino acid putative GTP exchanging factor related to RCC1.

RCC1 (regulator of chromosome condensation) is a eukaryotic protein which binds to chromatin and interacts with ran, a nuclear GTP-binding protein. RCC1 promotes the exchange of bound GDP with GTP, acting as a guanine-nucleotide dissociation stimulator.

The new protein can find application in the regulation of gene expression by activation of nuclear GTP-binding proteins. The X-linked retinitis pigmentosa is a result of a defect GTPase regulator, which contains a RCC1-type repeat.

similarity to RCC1-like G exchanging factor RLG

complete cDNA, complete cds, EST hits

Sequenced by LMU

Locus: /map="20"

Insert length: 2321 bp

Poly A stretch at pos. 2293, polyadenylation signal at pos. 2262

```

1 GGGTCACGCA AGATGGCGGC GCCCAGAGGC TGCTGAGGCG CGGAACGGAG
51 GATGGCGCTG GTGGCGTTGG TGGCTGGGGC TCGGCTGGGG CGGCGGCTGA
101 GCGGGCCGGG GCTGGGGCGA GGGCACTGGA CGGCGGCCAG GCGCTCCCGG
151 AGCCGGCCGG AAGCGGCAGA AGCCGAGGCG GAGGTGCCCG TGGTCCAGTA
201 CGTGGGCGAG CGCGCTGCCG GCGCCGATCG CGTCTTCGTG TGGGGCTTCA
251 GCTTCTCGGG GCGCTGGGC GTGCCTTCCT TTGTGGTGCC CAGCTCCGGG
301 CCCGGGCCCC GCGCGGGCGC CCGACCGCGC CGCAGGATCC AGCCCGTGCC
351 CTATCGCGCT GAGCTGGACC AAAAGATTTC ATCTGCTGCT TGCGGCTATG
401 GATTACACT GCTGTCTCTT AAGACTGCGG ATGTTACGAA AGTCTGGGGG
451 ATGGGACTCA ACAAAGATTC TCAGCTTGGA TTTACAGGA GCCGGAAGA
501 TAAACAGAGG GGCTACGAGT ATGTGTTGGA GCCCTCACCC GTCTCCCTGC
551 CTCTGGACAG ACCTCAGGAG ACACGGGTGC TGCAGGTCTC CTGCGGCCGA
601 GCTCACTCTC TTGTGTTGAC TGACAGGGA GGAGTCTTCA GCATGGGAAA
651 CAATCTTAT GGGCAATGTG GAAGAAAGGT GGTGCGAAAT GAAATTTACA
701 GTGAAAGTCA CAGAGTCCAC AGGATGCAGG ACTTCGATGG CCAGGTGGTC
751 CAGGTGCGCT GTGGTCAGGA TCATAGTCTG TTCCTGACGG ATAAAGGAGA
801 AGTCTATTCT TGTGGATGGG GTGCTGATGG GCAACAGGT CTGGGTCACT
851 ACAATATCAC CAGCTCGCCC ACCAAGCTGG GTGGAGACCT GGCGGGAGTG
901 AACGTTATCC AAGTTGCCAC CTACGGTGAT TGCTGCCTGG CCGTGTCCGC
951 CGACGGAGGA CTTTGTGGTT GGGGAACTC GGAGTACCTG CAGCTGGCCT
1001 CTGTCACTGA CTCCACACAG GTGAATGTGC CCCGCTGCTT ACCTTCTCA
1051 GGAGTGGGGA AGGTGCGACA GGCTGCATGC GGTGGCACGG GCTGTGCAGT
1101 GTTAAACGGA GAAGGACATG TTTTGTCTG GGGCTATGGA ATTCTGGGA
1151 AAGGTCCAAA CCTAGTGGA AGTGCCGTCC CTGAAATGAT TCCACCCACT
1201 CTCTTTGGCT TGACGGAGTT CAACCCAGAA ATCCAGGTTT CCCGCATCCG
1251 ATGTGGACTC AGCCACTTTG CTGCACTGAC CAACAAAGGA GAGCTGTTTG
1301 TATGGGGCAA GAACATCCGA GGTGCGCTGG GAATCGGTCG CCTGGAGGAC
1351 CAGTATTTC CATGGAGGGT GACGATGCCT GGGGAGCCTG TGGACGTGGC
1401 ATGTGGCGTG GACCACATGG TGACCCTGGC CAAGTCATTC ATCTAAACCT
1451 CCTCACCTG CTTGGGCGGC CCCGTCCCGG GAACCACTGG CACTCCTTGG
1501 CAGAGGCCAG CGCGTGGCCA GCCCCCCGGG GTTCTTGAT GGTGGTGGCG
1551 GAGGACCCTG CGTGCAGTGT GACGCTCTGT CCTGAATCCC TTAGCGGGTA
1601 CCTACCAGGA GGATCAGGGC AAGGTCCCTC TCCAGCTGCA GGTGAGGCCT
1651 GCGGAACCTA GCTTGGATGG CAGCCTTTGG TGGGCCGCTG TGGCCCGCAC
1701 GTCTCTGTTC TCTCCAAGTA ACATGCGACG GTGTCTGGTG TCACGTCTCG
1751 CCTGAGAAGC CCGTCTTAGG AAAGCTTAGC TTGAACACAG TGCTCGGGAG
1801 GTTCTGTGCT TGTCTGTCTT GGCAGTCTCT TGGTTTGTGT CTGGCCAAGG
1851 CCATGCGTGT GCCTCGGACC GAGCCCCAGC TTAGGCGAGG GAGTCAGGCT
1901 GGCCTCGGCC CTCGGTTTTC ATTCAGGCCA CCCTGCTCAT GGCCCTTCCT
1951 GGCCGCTGCG CACACCGCAA GCTCGCTGGG GGGACACTAG AAGCACCCTG
2001 GCCTGGGATT CCATCTGGAG CTGTCCGCAG GCACCAGCCC CAGCCTCCCA
2051 CCACGCTCAC TGCTGGCTT GGAAGGTTA AGAAGCCCTT CAGGAAGAGA
2101 ATCGAGGCTA AGTTCCTCTG CGCCGAGGGC CCCGAGCATA TCCGCCAAGG
2151 CTAGCTGCA GTGCCAGGCG GAGGAGGAAG ATCCAGAAAT TGTGAACAAT
2201 GTTTGATTTA GTAGCGTGAC TTGCCTTTCC CTTTAAAAAC ATCTTTTACA
2251 AATCTGTCTT GGAATAAAGT CTATTTTCTG CCTTTTGTTT TTTAAAAAAA
2301 AAAAAAAAAA AAAAAAAAAA A

```

BLAST Results

Entry HS203358 from database EMBL:
 human STS SHGC-31781.
 Score = 1748, P = 1.1e-72, identities = 376/394

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 52 bp to 1443 bp; peptide length: 464
 Category: similarity to known protein

```

1 MALVALVAGA RLGRRLSGPG LGRGHWTAAR RSRSRREAAE AEAEPVVPVQY
51 VGERAARADR VFVWGFSSFG ALGVPSFVVP SSGPGPRAGA RPRRRRIQVPV
101 YRLELDQKIS SAACGYGFTL LSSKTADVTK VWGMGLNKDS QLGFHRSRKD
151 KTRGYEYVLE PSPVSLPLDR PQETRVLQVS CGRAHSLVLT DREGVFMSGN
201 NSYGGQGRKV VENEIYSESH RVHRMQDFDG QVVQVACGQD HSLFLTDKGE
251 VYSCGWGADG QTGLGHYNIT SSPTKLGGDL AGVNVIVQVAT YGDCCLAVSA
301 DGGLEFGWNS EYLQLASVTD STQVNVPRCL HFSGVGKVRQ AACGGTGCAV
351 LNNEGHHVFVW GYGILGKGNP LVESAVPEMI PPTLFGLTEF NPEIQVSRIR
401 CGLSHFAALT NKGELFVWGK NIRGCLGIGR LEDQYFPWRV TMPGEPVDVA
451 CGVDHMTLA KSFI
  
```

BLASTP hits

Entry CEW09G3_5 from database TREMBLNEW:
 gene: "W09G3.3"; Caenorhabditis elegans cosmid W09G3
 Score = 395, P = 9.3e-37, identities = 111/330, positives = 165/330

Entry Y032_HUMAN from database SWISSPROT:
 HYPOTHETICAL PROTEIN KIAA0032.
 Score = 309, P = 1.0e-24, identities = 96/308, positives = 143/308

Entry B38919 from database PIR:
 hypothetical protein 2 - human (fragment)
 Score = 309, P = 1.0e-24, identities = 96/308, positives = 143/308

Entry AF060219_1 from database TREMBLNEW:
 product: "RCC1-like G exchanging factor RLG"; Homo sapiens RCC1-like G
 exchanging factor RLG mRNA, complete cds.
 Score = 273, P = 4.0e-21, identities = 84/262, positives = 124/262

Entry S71752 from database PIR:
 giant protein p619 - human
 Score = 282, P = 1.1e-19, identities = 86/287, positives = 144/287

Alert BLASTP hits for DKFZphtes3_21d4, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_21d4, frame 1

Report for DKFZphtes3_21d4.1

```

[LENGTH]      464
[MW]           49997.08
[pI]           8.74
[HOMOL]        TREMBL:CEW09G3_5 gene: "W09G3.3"; Caenorhabditis elegans cosmid W09G3 5e-34

[FUNCAT]       04.07 rna transport [S. cerevisiae, YGL097w] 2e-09
[FUNCAT]       03.07 pheromone response, mating-type determination, sex-specific proteins
               [S. cerevisiae, YGL097w] 2e-09
[FUNCAT]       08.01 nuclear transport [S. cerevisiae, YGL097w] 2e-09
[FUNCAT]       04.05.05 mrna processing (5'-end, 3'-end processing and mrna degradation) [S.
               cerevisiae, YGL097w] 2e-09
[FUNCAT]       04.01.04 rrna processing [S. cerevisiae, YGL097w] 2e-09
[FUNCAT]       04.03.03 trna processing [S. cerevisiae, YGL097w] 2e-09
[FUNCAT]       30.03 organization of cytoplasm [S. cerevisiae, YGL097w] 2e-09
  
```

```

[FUNCAT]      30.04 organization of cytoskeleton [S. cerevisiae, YAL020c] 4e-06
[BLOCKS]      BL008701
[BLOCKS]      BL00625B Regulator of chromosome condensation (RCC1) proteins
[BLOCKS]      BL00625A Regulator of chromosome condensation (RCC1) proteins
[PIRKW]       blocked amino end 3e-16
[PIRKW]       nucleus 3e-16
[PIRKW]       duplication 4e-08
[PIRKW]       tandem repeat 3e-16
[PIRKW]       DNA binding 3e-16
[PIRKW]       mitosis 3e-16
[PIRKW]       leucine zipper 3e-21
[SUPFAM]      pheromone response pathway component SRM1 4e-08
[SUPFAM]      WD repeat homology 3e-21
[PROSITE]     MYRISTYL 7
[PROSITE]     RCC1_2 2
[PROSITE]     AMIDATION 2
[PROSITE]     CAMP_PHOSPHO_SITE 1
[PROSITE]     CK2_PHOSPHO_SITE 5
[PROSITE]     TYR_PHOSPHO_SITE 2
[PROSITE]     GLYCOSAMINOGLYCAN 3
[PROSITE]     PKC_PHOSPHO_SITE 7
[PROSITE]     ASN_GLYCOSYLATION 2
[PFAM]        Regulator of chromosome condensation (RCC1)
[KW]          All Beta
[KW]          LOW_COMPLEXITY 13.58 %

```

```

SEQ  MALVALVAGARLGRRLSGPGLGRGHWTAAARRSRSRREAAEAEVPPVQYVGERAARADR
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  ccchhhhhhhhhheccccccccchhhhhhhhhhhhhhhhhhhhhhhceeeeehhhhhhhhhh

```

```

SEQ  VFVWGFSGALGVPSFVVPSSGPGPRAGARPRRIQVPYRLELDQKISSAACGYGFTL
SEG  .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  eeeeeccccccccceeeeeccccccccccccccccccccccccchhhhhhhheeeccccceee

```

```

SEQ  LSSKTADVTKVWGMGLNKDSQLGFHRSRKDKTRGYEYVLEPSVSLPLDRPQETRVLQVS
SEG  .....
PRD  eccccceeeccccccccccccccccccccccccceeeccccccccccccccccceee

```

```

SEQ  CGRAHSLVLTDRGVFSMGNNSYQCGRKRKVENIYSESHRVHRMQDFDQVQVQVACGQD
SEG  .....
PRD  cccccceeeccccceeeccccccccccccccccccccccccccccccccccccceeecccccc

```

```

SEQ  HSLFLTDKGEVYSCGWADGQTGLGHYNITSSPTKLGGDLAGVNVQVATYGDCCCLAVSA
SEG  .....
PRD  eeeeeccccceeeccccccccccccccccccccccccccccccccccccceeeccccceeeec

```

```

SEQ  DGGLFGWGNSEYLQLASVTDSTQVNVPRCLHFSGVGKVRQAACGGTGC AVLNGEGHVFW
SEG  .....
PRD  cccccccccccccccccccccccccccccccccccccceeeccccceeeccccccccceee

```

```

SEQ  GYGILGKGNLVSAPPEMIPPTLFLTEFNPEIQVSRIRGLSHFAALTNGELFWGK
SEG  .....
PRD  cccccccccccccccccccccceeeccccceeeccccccccceeeccccceeecccc

```

```

SEQ  NIRGCLGIGRLEDQYFPWRVTMPGEPVDVACGVDMVTLAKSFI
SEG  .....
PRD  cccccccccccccccccceeeccccceeecccccccccccccc

```

Prosites for DKFZphtes3_21d4.1

PS00001	200->204	ASN_GLYCOSYLATION	PDOC00001
PS00001	268->272	ASN_GLYCOSYLATION	PDOC00001
PS00002	17->21	GLYCOSAMINOGLYCAN	PDOC00002
PS00002	82->86	GLYCOSAMINOGLYCAN	PDOC00002
PS00002	333->337	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	14->18	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	34->37	PKC_PHOSPHO_SITE	PDOC00005
PS00005	122->125	PKC_PHOSPHO_SITE	PDOC00005
PS00005	147->150	PKC_PHOSPHO_SITE	PDOC00005
PS00005	190->193	PKC_PHOSPHO_SITE	PDOC00005
PS00005	219->222	PKC_PHOSPHO_SITE	PDOC00005
PS00005	246->249	PKC_PHOSPHO_SITE	PDOC00005
PS00005	410->413	PKC_PHOSPHO_SITE	PDOC00005
PS00006	34->38	CK2_PHOSPHO_SITE	PDOC00006
PS00006	147->151	CK2_PHOSPHO_SITE	PDOC00006
PS00006	190->194	CK2_PHOSPHO_SITE	PDOC00006
PS00006	290->294	CK2_PHOSPHO_SITE	PDOC00006
PS00006	317->321	CK2_PHOSPHO_SITE	PDOC00006

PS00007	209->217	TYR_PHOSPHO_SITE	PDOC00007
PS00007	208->217	TYR_PHOSPHO_SITE	PDOC00007
PS00008	9->15	MYRISTYL	PDOC00008
PS00008	20->26	MYRISTYL	PDOC00008
PS00008	133->139	MYRISTYL	PDOC00008
PS00008	238->244	MYRISTYL	PDOC00008
PS00008	277->283	MYRISTYL	PDOC00008
PS00008	302->308	MYRISTYL	PDOC00008
PS00008	344->350	MYRISTYL	PDOC00008
PS00009	12->16	AMIDATION	PDOC00009
PS00009	206->210	AMIDATION	PDOC00009
PS00626	179->190	RCC1_2	PDOC00544
PS00626	235->246	RCC1_2	PDOC00544

Pfam for DKFZphtes3_21d4.1

HMM_NAME	Regulator of chromosome condensation (RCC1)		
HMM	*IAaGqHHTVCLTqDGRVYtWG*		
	+A GQ+H++ LT++G VY++G		
Query	235	VACGQDHSFLFLTDKGEVYSCG	255

DKF2phtes3_21j15

group: transcription factors

DKF2phtes3_21j15 encodes a novel 898 amino acid protein with similarity human NY-CO-33 protein.

NY-CO-33 is a protein recognised by autologous antibodies of human colon cancer patients. The novel protein contains 4 C2H2 Zinc fingers and is a new putativ transcription factor.

The new protein can find application in modulating/blocking the expression of genes controlled by this transcription factor.

strong similarity to "NY-CO-33"

complete cDNA, complete cds, potential start at bp 27, EST hits

Sequenced by LMU

Locus: unknown

Insert length: 4407 bp

Poly A stretch at pos. 4321, polyadenylation signal at pos. 4301

```

1 CGCTGCAGCA GGTGTCACAG AGCCGCATGC TCCCGGAGCC CAGCCTCTTC
51 AGCACCCGTGC AGCTGTACCG GCAGAGCAGC AAGCTCTATG GCTCCATCTT
101 CACGGGGGCC AGCAAGTTCC GCTGTAAGGA CTGCAGCGCT GCCTACGACA
151 CCCTGGTGGA GTTGACAGTG CACATGAACG AGACGGGGCA TTACCGCGAC
201 GACAACCATG AGACCGATAA CAACAACCCC AAGCGCTGGT CCAAGCCTCG
251 CAAACGCTCC TTGCTGGAAA TGGAAAGGAA GGAAGACGCC CAGAAGGTGC
301 TGAAGTGCAT GACTGTGGC CACTCCTTTG AGTCCCTGCA GGATTTGAGT
351 GTCCATATGA TCAAAACAAA AACTTACCAA AAAGTGCTTC TGAAGGAACC
401 CGTCACTCCT GTCGCCGCCA AAATCATCCC TGCCACTCGG AAGAAAGCTT
451 CCCTGGAGCT GGAGCTCCCC AGCTCCCCAG ATTCCACAGG TGGAAACCCC
501 AAAGCCACCA TCTCAGACAC CAACGATGCA CTTCAGAAGA ACTCCAACCC
551 TTACATCACG CCAATAATC GGTACGGCCA CCAGAATGGG GCCAGCTATG
601 CATGGCACTT TGAGGCCCGG AAGTCGCAGA TCCTGAAGTG CATGGAGTGT
651 GGGAGCTCGC ATGACACCCT GCAGGAGCTC ACTGCCACAC TGATGGTCAC
701 TGGCCACTTC ATCAAGGTCA CCAACTCTGC TATGAAAAGG GGGAGGCCCA
751 TTGTGGAGAC GCCTGTCACA CCTACCATCA CAACCCTGCT GGATGAGAAG
801 GTCCAGTCCG TGCCCTTGGC AGCCACCACC TTCACGTCCC CCTCCAATAC
851 ACCTGCCAGC ATCTCCCAA AACTGAATGT GGAGGTCAAG AAGGAAGTCG
901 ACAAGGAGAA AGCGGTCAC GACGAGAAAC CTAAGCAAAA AGACAAGCCT
951 GGGGAAGAAG AGGAGAAGTG TGACATCTCT TCCAAATACC ATTACTTGAC
1001 TGAAGATGAC TTAGAAGAGA GTCCCAAGGG GGGGCTTGAT ATCCTCAAAT
1051 CCTTGGGAAA CACAGTGACA TCCGCAATCA ACAAGGCCCA GAACGGCACT
1101 CCTAGCTGGG GGGGCTATCC CAGCATCCAT GCCGCTTACC AACTTCCCAA
1151 CATGATGAAG TTGTCCCTGG GCTCGTCCGG GAAGAGCAGC CCCCTGAARC
1201 CCATGTTTGG CAACAGTGAG ATTGTCTCCC CGACGAAAAA CCAGACCTTG
1251 GTCTCTCCAC CCAGCAGCCA GACGTCCCCC ATGCCCAAGA CAACTTTTCA
1301 TCCCATGGAG GAGCTGGTGA AAAAGGTAC TGAGAAAGTT GCCAAAGTGG
1351 AGGAGAAGAT GAAGGAGCCG GATGGGAAGC TTTCCCGGCC CAAGCGGGGCC
1401 ACTCCCTCCC CATGTAGCAG CGAAGTCGGG GAACCCATCA AGATGGAGGC
1451 ATCCAGCGAT GGGGGCTTCC GCAGCCAGGA GAACAGCCCC AGCCCCCGCG
1501 GGGATGGGTG CAAGGATGGG AGCCCCCTCG CTGAGCCGGT GGAGAATGGC
1551 AAGGAGCTGG TGAAGCCCTT AGCCAGCAGT TTGAGTGGCA GCACGGCCAT
1601 CATCACCGAC CACCCGCCCTG AACAGCCTTT TGTTAACCTT TTGAGCGCCC
1651 TGCACTCAGT CATGAACATT CACCTGGGCA AGGCCGCCAA GCCCTCCCTG
1701 CCTGCCCTGG ACCCCATGAG CATGCTTTTC AAGATGAGCA ACAGCCTGGC
1751 GGAGAAGGCT GCTGTGGCCA CCCC GCCGCC CCTGCAGTCC AAGAAGGCAG
1801 ACCACCTCGA CCGCTATTTC TACCAGTCA ACAACGACCA GCCCATAGAC
1851 TTGACAAAAG GGAAGAGTGA CAAAGGCTGC TCCTTGGGTT CAGTGCTTCT
1901 GTCACCCACG TCCACAGCCC CGGCAACCTC CTCATCCACG GTGACAACGG
1951 CAAAGACATC TGCCGTCTGT TCATTCTATG CAAACTCGCC GCTACGCGAG
2001 AATGCCCTTG CAGATATATC CGATATGCTG AAGAACTTGA CAGAGAGCCA
2051 CACGTCAAAA TCCTCCACTC CTTCCAGCAT CTCGAGAAAG TCTGACATTG
2101 ACGGGGCCAC TCTGGAGGAG GCTGAGGAGT CGACGCCCGC CCAGAAGAGG
2151 AAGGGCCGCC AGTCAAACTG GAACCCCGAG CACCTCTGTA TCCTCCAGGC
2201 CCAGTTTGCC GCCAGCCTCC GGCAGACCTC AGAAGGGAAG TACATCATGT
2251 CAGACCTGAG CCCCAGGAG CGGATGCATA TCTCCAGGTT CACCGGGCTG
2301 TCCATGACCA CCATCAGCCA CTGGCTGGCC AACGTGAAAT ACCAGCTTGG
2351 AAGGACAGGT GGAACAAAGT TCCTCAAAA CTTGGACACT GGCCACCCCG
2401 TCTTCTTTTG TAACGATTGT GCGTCCCAAA TCAGGACTCC TTCCACGTAC
2451 ATCAGTCACC TAGAGTCACA CTTAGGCTTC CGGCTACGGG ACTTATCCAA
2501 ACTGTCCACC GAACAGATTA ACAGTCAGAT AGCACAAACC AAGTCACCGT
2551 CAGAAAAAAT GGTGACGTCC TCCCCGAGG AAGACCTGGG GACTTCCCTAT
2601 CAGTGCAAAAC TTTGCAATCG GACCTTTGCC AGCAAGCAGC CTGTTAAACT

```

```

2651 TCACCTTAGC AAAACACACG GGAATCTCC GGAAGACCAC CTTCTGTATG
2701 TCTCTGAGTT AGAGAAGCAG TAGCATTTCG TTTTGATAGA AAGGACTGCA
2751 GTTTGCTTTG AGGGAACCTG TGGAAAGGCAC CTTCAGGCCC CCTCTGACTT
2801 GTTGTCTCTG GCACATGTTT TTATTTTAAC TGCAGAGAAT CACTCTGGGC
2851 TGGACTGTTT TGTATAACTG TACAGTGTTC AATAGAGGTG CATATCAGCG
2901 TGTGTGTTACT GGTAAAATAT GAAGGTTAAA ATGCAGTGGT AAGTGTGTTG
2951 AACTTTGTGT AAACGGGATT TAGTTGTGAG CATCCTCCCG ATGCTTCAAG
3001 CTGCATGCAT TAACAGACAG TTTAATTAAG CATTATAAAC GGAATCAGGC
3051 ACACCTTTTC CACGAGACTC GAGTGTGCTG GCATTCTCA CCCTTCATC
3101 TTTAGCCCTC TGAGTACTTT GAAGCACTTT TGCATTAAAT TGGTTAAAAA
3151 ATAAAAATAA ATAATAATAA TGTATGAAGC TCTGTTTTTT AAACCTCTTA
3201 CCAGCTTAGT TATAATGAAT AATATGAACC TCCATTTATG CAGGTCTGCA
3251 GGGGTATAAC ACGCCTTGAA ATTTAAAAGA ATATTATTTT CACATTGAAA
3301 CATAGATGTA TATATTGTAT AGATTTCAGA CTCTCTTATG AAAAAAATG
3351 TGATTGTGGT TAAATGACCT TTTCTTGCA TTTATAGCAA CAGTGTTTTA
3401 TGCACCTGCT ATGCTCTGGG CATAAGCTGT GCCTATGTAT AGTGTATATT
3451 TCTTTTTTTC TTTTTTTTAA GGTCTATGGG TTTTGTGTTT TACATGCAAA
3501 CATGTGTAAT TATACAGAAG ATACCACAGA TAGCATTAT AAAGTATACA
3551 GAAACATTAT CTGAAAGCAA AGTATGATAG TTTGTTTTGC TATACAGTAC
3601 ATCTATATTG ATAGAGGTTT ATGTTTAAAT TATACATATT TATTAGCATC
3651 ATATTGTGAT TTGTTTTGAG CAGTCTGAAT AAACGAGACC GGGAAAGACA
3701 TCCCTGGCAG GCATCAGAAC TATTTGACAC ATGATTTTAA AAGGTATTTA
3751 TTAGAAATCA AAGAACACTC AAAATAAACT CAGTGTCTAA AGGTTAAGT
3801 CTATTTGAAA AGGTTAAAAA AAAGAACAAA AAAAAAATAA GAACCTGTAC
3851 TGTATTTCTT AAACATTGAT AAAGCCTTTA AAATGTTTGT ACTGTAATAC
3901 TTTGCTTAAA AGTCATGAGG CATTCTGTGA TCCAACCTCT TTCACTTATT
3951 TATAAGCCCT CTTGGTTGCT ATTCCATATT GTAGGATGCC TTTCTATTTC
4001 AATTGGTAAC TTTCTGTTTT GTTCTTCTTA ATTATTCTCC CAAGATCCCA
4051 CACTGCAGCT TTATCTTTAG GCTTATGAAA GGTAAACCGT GGTACCAGGC
4101 TCTCCAAGTG ATTCTGTTCT TCTCCATTTT TGGCAGTTAA TTTGCAGAAG
4151 TAACTGACAG CTGACACCAT ATGAGAACCT TTGTATAAAA TATTGGCATG
4201 TAAACAGCAG AGACACCGTA ACACACTCTG TGCCCTGTTT GGTGTTGAC
4251 AATGAAGCAC CATTATGTGA CTCTTCATAT AACCTTTTTT TCTACGGCAG
4301 CATTAATAAT GTCTTTTTCG TATAAAAAAA AAAAAAATAA AAAAAAATAA
4351 AAAAAAATAA AAAAAAATAA AAAAAAATAA AAAAAAATAA AAAAAAATAA
4401 AAAAAAATAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 27 bp to 2720 bp; peptide length: 898
 Category: strong similarity to known protein

```

1 MLPEPSLFST VQLYRQSSKL YGSIFTGASK FRCKDCSAAY DTLVELTVHM
51 NETGHYRDDN HETDNNNPKR WSKPRKRSLL EMEGKEDAQK VLKCMYCGHS
101 FESLQDLSVH MIKTKHYQKV PLKEPVTTPVA AKIIPATRKK ASLELELPSS
151 PDSTGGTPEKA TISDTNDALQ KNSNPYITPN NRYGHQNGAS YAWHFEEKRS
201 QILKCMCEGS SHDTLQELTA HMMVTGHHFIK VTNSAMKKGK PIVETPVTPT
251 ITTLLDEKVQ SVPLAATFTT SPSNTPASIS PKLNVEVKKE VDKEKAVTDE
301 KPKQKDKPGE EEEKCDISSK YHYLTENDLE ESPKGGLDIL KSLNTVTSA
351 INKAQNGTPS WGGYPSIHAA YQLPNMMKLS LGSSGKSTPL KPMFGNSEIV
401 SPTKNQTLVS PPSQTSPPMP KTNFHAMEEL VKKVTEKVAK VEEKMKEPDG
451 KLSPPKRATP SPCSSEVGEP IKMEASSDGG FRSQNSPSP PRDGCKDGGSP
501 LAEPVENGKE LVKPLASSLS GSTAIITDHP PEQPFVNPLS ALQSVMNHL
551 GKAAKPSLPA LDPMSMLFKM SNLSAEKAAV ATPPLQSKK ADHLDRYFYH
601 VNNDQPIDLT KGKSDKGCST GSVLLSPTST APATSSSTVT TAKTSAVVSF
651 MSNSPLRENA LSDISDMLKN LTESHTSKSS TPSSIIEKSD IDGATLEEAE
701 ESTPAQKRKG RQSNWNQOHL LILQAQFAAS LRQTSSEKGI MSDLSPOERM
751 HISRTGLSM TTISHWLANV KYQLRRTGGT KFLKNLDTGH PVFFCNDACS
801 QIRTPSTYIS HLESHLGFRD RDLSKLSTEQ INSQIAQTKS PSEKMTVSSP
851 EEDLGTSYQC KLCNRTFASK HAVKLHLSKT HGKSPEDHLL YVSELEKQ

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_21j15, frame 3

TREMBL:AF039698_1 gene: "NY-CO-33"; product: "antigen NY-CO-33"; Homo sapiens antigen NY-CO-33 (NY-CO-33) mRNA, complete cds., N = 1, Score = 1039, P = 5.5e-105

PIR:A38437 probable homeotic protein tsh - fruit fly (Drosophila melanogaster), N = 3, Score = 158, P = 7.2e-09

TREMBL:CE33058_1 gene: "unc-89"; product: "UNC-89"; Caenorhabditis elegans UNC-89 (unc-89) gene, complete cds., N = 2, Score = 175, P = 3.3e-07

>TREMBL:AF039698_1 gene: "NY-CO-33"; product: "antigen NY-CO-33"; Homo sapiens antigen NY-CO-33 (NY-CO-33) mRNA, complete cds.
Length = 687

HSPs:

Score = 1039 (155.9 bits), Expect = 5.5e-105, P = 5.5e-105
Identities = 244/504 (48%), Positives = 319/504 (63%)

Query: 170 QKNSNPYITPNNRYGHQNGASYAWHFEARKSQILKCMCGSSHDTLQELTAHMMVTGHFI 229
QK +NPY+TPNNRYG+QNGASY W FEARK+QILKCMCGSSHDTLQ+LTAHMMVTGHF+
Sbjct: 14 QKAANFYVTPNNRYGQNGASYTWQFEARKAQILKCMCGSSHDTLQQLTAHMMVTGHFL 73

Query: 230 KVTNSAMKKGKPIVETPVTPTITLLDEKQSVPLAATFTS-PSNT---PASISPKLN 284
KVT SA KKGK +V PV ++EK+QS+PL TT T P+++ P S +
Sbjct: 74 KVTTSASKKGKQLVLDPV-----VEEKIQSIPLPPTHTRLPASSIKKQPDSPAGSTT 126

Query: 285 VEVKKEVDKEKA-VTDEKPKQKDKPGEEEEKCDISSKYHYLTENDLEESPKGGDLILKSL 343
E KKE +KEK V + K K++ + EK + S+ Y YL E DL++SPKGGDLILKSL
Sbjct: 127 SEEKKEPEKEKPPVAGDAEKIKEESDSELEKFEPTSTLYPYLREEDLDDSPKGGDLILKSL 186

Query: 344 ENTVTSAINKAQNGTPSWGYPYPSIHAAYQLPNNMKLSLGSSTKSTPLKPMF-GNSEIVSP 402
ENTV++AI+KAQNG PSWGGYPYPSIHAAYQLP +K L ++ +S ++P + G + +S
Sbjct: 187 ENTVSTAISKAQNGAPSWGYPYPSIHAAYQLPGTVK-PLPAAVQSVQVQPSYAGGVKSLSS 245

Query: 403 TKNQTLVSPSSQTSMPKTNFAMEELVKKVTEKV-AKVEEKMKEPDGKLSPPKRAPTS 461
++ L+ P S T P K+N AMEELV+KVT KV K EE+ E + K S K A S
Sbjct: 246 AEHNALLHSPGSLTPPHKSNVSAMEELVEKVTGKVNKKEERPPPEKE-KSSLAKAA--S 302

Query: 462 PCSSEVGEPKMEASSDGGFRSQENSPSPPRDGCKDGSPLAEPVENGKELVKPLASSLSG 521
P + E + K E S + Q+ P K PL NG E +K ++
Sbjct: 303 PIAKENKDFPKTEEVSG---KPQKKGPEAETWEAKKEGPLDVHTPNGTEPLKAKVTNGCN 359

Query: 522 STAIITDHPPEQPFVNPLSALQSVMMNIHLGKAAKPSLPALDPMMLFKMSNSLAEKAABA 581
+ II DH PE F+NPLSALQS+MN HLKG +KP P+LDP++ML+K+SNS+ +K
Sbjct: 360 NLGIIMDHSPEPSFINPLSALQSIMNTHLGKVSFVSPSLDPLAMLYKISNSMLDKPVYP 419

Query: 582 TPPPLQSKKADHLDYFYHVNNDQPIDLTGKGSCK-GCSLGSVLLSPTSTAPATSSSTVT 640
P K+AD +DRY+Y N+DQPIDLTGK+ S+ + SP + S +
Sbjct: 420 ATPV---KQADAIDRYYYE-NSDQPIDLTGKSNKPLVSSVADSVASPLRESALMDISDMV 475

Query: 641 TAKTSAVVSFMSN-SPLRENALSDISDMLKNL 673
T + S S + E + +D S + L E
Sbjct: 476 KNLTGRLTPKSSTPSTVSEKSDADGSSFEALDE 509

Score = 865 (129.8 bits), Expect = 7.4e-95, P = 7.4e-95
Identities = 211/434 (48%), Positives = 268/434 (61%)

Query: 447 EPDGKLSPPKRAPTSPPCSSEVG--EPIKMEASSDGGFRSQENSPSPPRDG-CKDGSPLAE 503
E + L P T P P S V E + + + + +E P + K SP+A+
Sbjct: 247 EHNALLHSPGSLTPPHKSNVSAMEELVEKVTGKVNKKEERPPPEKEKSSLAKAASPIAK 306

Query: 504 -----P-VE--NGKELVK-PLASSLSGSTAIITD-HPPE--QPFVNPLSALQSVMMNIHLG 551
P E +GK K P A + D H P +P ++ + + I +
Sbjct: 307 ENKDFPKTEEVSGKPKQKKGPEAETWEAKKEGPLDVHTPNGTEPLKAKVTNGCNLGIIMD 366

Query: 552 KAAKPSLPALDPMMLFKMSNSLAEKAABATPPPLQSKKADHLDYFYHVNNDQPID 608
+ +PS ++P+S L + N+ K + P L D L Y ++N D+P+
Sbjct: 367 HSPEPSF--INPLSALQSIMNTHLGKVSFVSPSL-----DPL-AMLYKISNSMLDKPV- 417

Query: 609 LTKGKSDKGCSLGSVLLSPTSTAPATSSSTVTTAKTSAVVSFMSNSPLRENALSDISDML 668
K S P + + S+V ++ SPLRE+AL DISDM+
Sbjct: 418 -YPATPVKQADAIDRYYYENSQPIDLTGKSNKPLVSSVADSVASPLRESALMDISDMV 475

Query: 669 KNLTESHTSKSSTPSSISEKSDIDGATLEEA-EESTPAQKRGRQSNWNPNQHLILQAQF 727
KNLT T KSSTPS++SEKSD DG++ EEA +E +P KRGGRQSNWNPNQHLILQAQF

Score = 98 (14.7 bits), Expect = 7.4e-95, P = 7.4e-95
Identities = 32/95 (33%), Positives = 47/95 (49%)

Report for DKFZphtes3_21j15.3

SEQ	MLPEPSLFSTVQLYRQSSKLYGSIFTGASKFRCKDCSAAYDTLVELTVHMNETGHYRDDN
SEG
PRD	ccccceeeeeeeccccceeeeeccccceeecccchhhhhhhhhhhcccccccccccc
SEQ	HETDNNNPKRWSKPKRKSLLMEGKEDAQKVLKCMYCGHSFESLQDLSVHMIKTKHYQKV
SEG
PRD	ccccccccccccccccchhhhhhhccchhhhhhhhhccccchhhhhheeeeeeeccceee
SEQ	PLKEPVTVPAAKIIIPATRKKASLELELPSSPDSTGGTPKATISDTNDALQKNSNPYITPN
SEGxxxxxxxxxx.....
PRD	ecccccceeeeeehhhhhhhhhhhccccccccccccccccceeeccchhhhhcccccccc
SEQ	NRYGHONGASYAWHFEARKSQILKCMCEGSSHDTLQELTAHMMVTGHFIKVTNSAMKKGK
SEG
PRD	ccccccccchhhhhhhhhhhhhhhhhccccccccchhhhhhhhhhhceeeeecccccccc
SEQ	PIVETPVTPTITTLDEKVSQVPLAATTFTSPNSNTPASISPKLNVEVKKEVDKEKAVTDE
SEG	xxxxxxxxxxxxxx.....xxxxxxxxxxxxxxxxxxxx
PRD	ccccccccccchhhhhhhheccccccccccccccccccccccccccccccccchhhhhhcc
SEQ	KPKOKDKPGEEEKCDISSKYHYLTENDLEESPKGGLDILKSLENTVTSAINKAONGTPS


```

SEG      x.....
PRD      cccccccccccccchhhhhhhhhccccccccchhhhhhhhhhhhhhhhhcccccc

SEQ      WGGYPSIAAAYQLPNMMKLSLGSSGKSTPLKPMFGNSEIVSPTKNQTLVSPSSQTSMP
SEG      .....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ      KTNFHAMEELVKVTEKVAKEEKMKEPDGKLSPPKRATPSPCSSEVGEPKMEASSDGG
SEG      ..... xxxxxxxxxxxxxxxxxxxxxxxx .....
PRD      ccchhhhhhhhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccc

SEQ      FRSQENSPSPRDGCKDGSPLAEPVENGKELVKPLASSLSGSTAIITDHPPEQPFVNPLS
SEG      .....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ      ALQSVMNHLGKAAKPSLPALDPMSMLFKMSNSLAEKAAVATPPPQLQSKADHLDRYFYH
SEG      .....
PRD      chhhhhccccccccccccchhhhhhhhhhhhhcccccccccccccccccccccccccc

SEQ      VNNDQPIDLTKGSKDKGCSLGSVLLSPTSTAPATSSSTVTAKTSAVVFSFMSNSPLRENA
SEG      ..... xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx .....
PRD      eccccceccccccccccccccccccccccccccccccccccccccccccccccccchh

SEQ      LSDISDMLKNLTESHTSKSSTPSSIIEKSDIDGATLEAEESTPAQKRKGRQSNWNPQHL
SEG      ..... xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx .....
PRD      hhhhhhhhhhhhhccccccccccccccccccccchhhhhhhccccchhhhhccccccccch

SEQ      LILQAQFAASLRQTSEGKYIMSDLSPQERMHISRTGLSMTTISHWLANVKYQLRRTGGT
SEG      .....
PRD      hhhhhhhhhhhhhccccccccccccchhhhhhhccccchhhhhhhhhhhhhhhcccc

SEQ      KFLKNLDTGHPVFFCNDCASQIRTPSTYISHLESHLGFRLRDL SKLSTEQINSQIAQTKS
SEG      .....
PRD      ceeccccccccccccccccccccchhhhhhhhhhhhhhhhhhhccchhhhhhhhhhhcc

SEQ      PSEKMVTSSPEEDLGTSYQCKLCNRTFASKHAVKLHLSKTHGKSPEDHLLYVSELEKQ
SEG      .....
PRD      cccccccccccccccccchhhhhhhhhhhhhhhhhhhcccccccccccccccccccccc

```

Prosites for DKFZphtes3_21j15.3

PS00001	51->55	ASN_GLYCOSYLATION	PDOC00001
PS00001	405->409	ASN_GLYCOSYLATION	PDOC00001
PS00001	670->674	ASN_GLYCOSYLATION	PDOC00001
PS00001	864->868	ASN_GLYCOSYLATION	PDOC00001
PS00004	69->73	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	75->79	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	139->143	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	432->436	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	456->460	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	17->20	PKC_PHOSPHO_SITE	PDOC00005
PS00005	137->140	PKC_PHOSPHO_SITE	PDOC00005
PS00005	157->160	PKC_PHOSPHO_SITE	PDOC00005
PS00005	280->283	PKC_PHOSPHO_SITE	PDOC00005
PS00005	318->321	PKC_PHOSPHO_SITE	PDOC00005
PS00005	332->335	PKC_PHOSPHO_SITE	PDOC00005
PS00005	384->387	PKC_PHOSPHO_SITE	PDOC00005
PS00005	435->438	PKC_PHOSPHO_SITE	PDOC00005
PS00005	588->591	PKC_PHOSPHO_SITE	PDOC00005
PS00005	614->617	PKC_PHOSPHO_SITE	PDOC00005
PS00005	641->644	PKC_PHOSPHO_SITE	PDOC00005
PS00005	676->679	PKC_PHOSPHO_SITE	PDOC00005
PS00005	686->689	PKC_PHOSPHO_SITE	PDOC00005
PS00005	730->733	PKC_PHOSPHO_SITE	PDOC00005
PS00005	842->845	PKC_PHOSPHO_SITE	PDOC00005
PS00006	42->46	CK2_PHOSPHO_SITE	PDOC00006
PS00006	78->82	CK2_PHOSPHO_SITE	PDOC00006
PS00006	103->107	CK2_PHOSPHO_SITE	PDOC00006
PS00006	149->153	CK2_PHOSPHO_SITE	PDOC00006
PS00006	161->165	CK2_PHOSPHO_SITE	PDOC00006
PS00006	210->214	CK2_PHOSPHO_SITE	PDOC00006
PS00006	214->218	CK2_PHOSPHO_SITE	PDOC00006
PS00006	253->257	CK2_PHOSPHO_SITE	PDOC00006
PS00006	325->329	CK2_PHOSPHO_SITE	PDOC00006
PS00006	573->577	CK2_PHOSPHO_SITE	PDOC00006
PS00006	684->688	CK2_PHOSPHO_SITE	PDOC00006
PS00006	689->693	CK2_PHOSPHO_SITE	PDOC00006
PS00006	695->699	CK2_PHOSPHO_SITE	PDOC00006
PS00006	745->749	CK2_PHOSPHO_SITE	PDOC00006

PS00006	810->814	CK2_PHOSPHO_SITE	PDOC00006
PS00006	840->844	CK2_PHOSPHO_SITE	PDOC00006
PS00006	848->852	CK2_PHOSPHO_SITE	PDOC00006
PS00006	884->888	CK2_PHOSPHO_SITE	PDOC00006
PS00006	893->897	CK2_PHOSPHO_SITE	PDOC00006
PS00007	732->740	TYR_PHOSPHO_SITE	PDOC00007
PS00007	883->892	TYR_PHOSPHO_SITE	PDOC00007
PS00008	22->28	MYRISTYL	PDOC00008
PS00008	156->162	MYRISTYL	PDOC00008
PS00008	188->194	MYRISTYL	PDOC00008
PS00008	362->368	MYRISTYL	PDOC00008
PS00008	479->485	MYRISTYL	PDOC00008
PS00008	494->500	MYRISTYL	PDOC00008
PS00008	498->504	MYRISTYL	PDOC00008
PS00008	617->623	MYRISTYL	PDOC00008
PS00008	757->763	MYRISTYL	PDOC00008
PS00028	795->816	ZINC_FINGER_C2H2	PDOC00028
PS00028	860->882	ZINC_FINGER_C2H2	PDOC00028
PS00028	33->56	ZINC_FINGER_C2H2	PDOC00028
PS00028	94->117	ZINC_FINGER_C2H2	PDOC00028

Pfam for DKFZphtes3_21j15.3

HMM_NAME Zinc finger, C2H2 type

HMM *CpwPDCgKtFrwsNLrRHMR..T.H*
 C++ C ++ + +L+ HM+ H

Query 33 CKD--CSAAYDTLVELTVHMMNET-GH 55

26.69 (bits) f: 94 t: 116 Target: dkfzphtes3_21j15.3 strong similarity to "NY-CO-33"
 Alignment to HMM consensus:

Query *CpwPDCgKtFrwsNLrRHMR..T.H*
 C + CG +F + +L HM+ H
 dkfzphtes3 94 CMY--CGHSFESLQDLSVHMIKT-KH 116

Query f: 795 t: 815 Target: dkfzphtes3_21j15.3 strong similarity to "NY-CO-33"
 Alignment to HMM consensus:

HMM *CpwPDCgKtFrwsNLrRHMRT*
 C++ C R++S+++ H+ +H
 Query 795 CND--CASQIRTPSTYISHLESH 815

27.12 (bits) f: 860 t: 881 Target: dkfzphtes3_21j15.3 strong similarity to "NY-CO-33"
 Alignment to HMM consensus:

Query *CpwPDCgKtFrwsNLrRHMR.T.H*
 C+ C++TF +++ + H+ H
 dkfzphtes3 860 CKL--CNRTFASKHAVKLHLSK-TH 881

DKFZphtes3_21116

group: intracellular transport and trafficking

DKFZphtes3_21116 encodes a novel 66 amino acid protein nearly identical to rat ribosome attached membrane protein 4 (ramp4).

The novel protein seems to be the human orthologue of rat ramp 4. Ramp4 is involved in the regulation of translocation of proteins into endoplasmic reticulum, e.g. of the MHC class II associated invariant (gamma) chain.

The new protein can find application in modulation of protein translocation into the endoplasmic reticulum.

identical to rat ribosome attached membrane protein 4

ORF Bp 316-513 (66 aa) see BLASTX

Sequenced by LMU

Locus: unknown

Insert length: 2488 bp

Poly A stretch at pos. 2464, polyadenylation signal at pos. 2442

```
1  CTTCTCTTTT  CACTCCGCGC  TCACGGCGGC  GGCCAAAGCG  GCGGCGACGG
51  CGGCGCGAGA  ACGACCCGCG  GGCCAGTTCT  CTTCTCCTTG  CGCACCTGCC
101 CCGCTCGGTC  AGTCAGTCGG  CGGCCGGCGC  CCGGCTTGTC  CTCAGACCTC
151 GCGCTTGCGG  CGCCCAGGCC  CAGCGGCCGT  AGCTAGCGTC  TGGCCTGAGA
201 ACCTCGGCGC  TCCGGCGGCG  CGGGCACCAC  GAGCCGAGCC  TCGCAGCGGC
251 TCCAGAGGAG  GCAGGCGAGT  GAGCGAGTCC  GAGGGGTGGC  CGGGGCAGGT
301 GGTGGCGCGC  CGAAGATGGT  CGCCAAGCAA  AGGATCCGTA  TGGCCAACGA
351 GAAGCACAGC  AAGAACATCA  CCCAGCGCGG  CAACGTCGCC  AAGACCTCGA
401 GAAATGCCCC  CGAAGAGAAG  GCGTCTGTAG  GACCCGTGGT  ATTGGCTCTC
451 TTCATTTTGT  TTGTCTGTGG  TTCTGCAATT  TTCCAGATTA  TTCAAAGTAT
501 CAGGATGGGC  ATGTGAAGTG  ACTGACCTTA  AGATGTTTCC  ATTCTCCTGT
551 GAATTTTAAC  TTGAACCTCAT  TCCTGATGTT  TGATACCCTG  GTTGAAAACA
601 ATTCAGTAAA  GCATCCTGCC  TCAGAATGAC  TTCTCTATCA  TGCTTCATGT
651 GTCATTCCAA  GGTTCCTTCA  TGAGTCATTC  CAAGTTTTCT  AGTCCATACC
701 ACAGTGCCTT  GCAAAAAACA  CCACATGAAT  AAAGCAATAA  AATTTGATTG
751 TTAAGATACA  GTAGTGGACC  CTACTTATTC  AGTCAATTAA  GAGTAAGTTT
801 TTTTATGTGG  TTATTTAAAC  AGTATGAACA  ATTAGTCTAA  CTCTGCATAG
851 ACAGGGTCTG  GATTTTGTTA  ACCCAAATGT  ATAAGTGCAG  TTAGCTTAAA
901 TTACAAATTT  AAGTCTTGTC  GTTTTATAT  AGCTAGGCAC  TTTATTAECT
951 TTTTGAACCT  AAAGCACACT  CCCTTATAGG  TTCATGTAAC  TGTCCTGTAA
1001 TAAGGTGCTT  ATAAATGGAA  CAACTACACA  GCCTAGTTTT  GCCACAACCT
1051 TTAGCATCTA  AAAAGTTTTA  AAAGCTCTTA  AATGTCTAAT  ATAAAGGGAG
1101 ATGCTTATAG  CCACACATC  TATTTTACCA  ATATTGTTTC  CATTACACTA
1151 CCTTGGATTT  TGCATGAGTG  AGTATAGTAA  CCCAAGATGC  CATAAAAAAA
1201 AACTTGATCG  TTTTCTGACT  TAATTAGTTA  CTGTGGTTTC  ACTAAAAGCT
1251 ACCGTGGTGG  AGTGAAGTCA  GTCAGGGAAG  GTTTGTTTAT  GTTACATTTA
1301 TTCAACAGAG  ACTATTTTAA  TATATCAAAG  GGGTTTACTA  TGCCAAACAA
1351 AATTCTAGGG  AAAAAATCTG  CTAATAATGG  ATGCCTCATC  AGAACATGCT
1401 GTTGAGTCCA  ATGTGCCATA  AGACATTTTA  GCATGTATAA  TAGCACTTTT
1451 AATAGCAAAA  AAAGGCACAT  CAACTGCGAA  GTTATCCTTA  GTTTGCAAAAT
1501 GCTTTTCTTA  GATTAATGAT  TTTTCAATCA  TTAGGGTACT  AGACACATCA
1551 GCCTAAAGTG  GCATCTGGAA  TTGAATGGAT  TTAGTGATAA  TGATCAGTCT
1601 TTAGTCTTCC  CTTTGTTATA  TGACTTTATA  GGTTATGATT  GATCAAAATT
1651 ACGTTTTACT  AATGGTAAGG  GTGAGGTC  TAGGGCAGGT  TTTGGGTTTT
1701 CTAGTACTGT  TGAAAACCTG  AAGTATTGGC  TATTTGTATA  CTTAGCCATA
1751 ACTTGGTGAA  AAAAAACCTG  AGCAGTGTCT  ATGTATTAAT  GCGTTGGAAA
1801 GAAAGCTGCT  TGTGTTTGCT  TTGTTAATTG  CCTCAGGATA  TTTCTTTTAA
1851 AATAAGCTGT  TTAAAGAGGA  ACAGAAGGGA  AATCTGCTAC  CTAGTCTATA
1901 CACAGCGTGA  ACCTCACAGG  GGGCTTCTGA  TACCCTCAAA  CATGGAGAAC
1951 AGTAAGGGAG  CAGAGTGGTT  AAGGACTTTC  AGGAACCTTA  CTATTCTGGA
2001 ATAAGGAATG  AATCAACTGA  CCTTGGGCCA  GCAGGTTTTT  AACTAAATTG
2051 TTACTTGCCT  TTCTCACCCA  GTTAATCAGT  CTCTGTACTT  GTTTCCTTTT
2101 TTGAAACAAG  TGTCTTGGTT  AACTAATTCT  GTTTATGGT  TGTGCTAAAT
2151 TCATAGCAGG  TGCCTTATTC  TTTGCTTTTA  GTCAAACCAT  TCCATATCAG
2201 AATTTTCTTT  GGTTTACTAT  AGATATTTGG  CTTTAAAGTG  TTGTTTGTGT
2251 TTTTAAATGT  ACAATGTTCT  GATAAATTTG  ACTGTTAAAT  TGCTATAGCT
2301 AGCAATCATT  TTACATATGT  AAAAAATTGC  ATTCCCTTTG  TATTTTCATG
2351 GTAATTCACC  AATTAAGTGC  AGTTTATATT  CAGGTTGGAT  TATGCATGTT
2401 TAGGTAAACG  AAAGCTGTGT  CTTACTTGAT  TTATTCTTTA  AAAATAAAGT
2451 TCCCTGAATA  TTTGAAAAAA  AAAAAAATAA  AAAAAAATAA
```

BLAST Results

Entry HSCDN13 from database EMBL:
H.sapiens (TL5) mRNA from LNCaP cell line
Score = 1075, P = 5.8e-41, identities = 219/221

Entry AF100470_1 from database TREMBLNEW:
gene: "RAMP4"; product: "ribosome attached membrane protein 4"; Rattus norvegicus ribosome attached membrane protein 4 (RAMP4) mRNA, complete cds.
Score = 331, P = 3.9e-28, identities = 66/66, positives = 66/66, frame +1

Entry HSG19910 from database EMBL:
human STS A002B48.
Score = 530, P = 2.1e-17, identities = 108/109

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 316 bp to 513 bp; peptide length: 66
Category: strong similarity to known protein
Classification: Intracellular transport and traffic

1 MVAKQIRIRMA NEKHSKNITQ RGNVAKTSRN APEEKASVGP WLLALFIFVV
51 CGSAIFQIIQ SIRMGM

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_21116, frame 1

TREMBLNEW:RNO238236_1 gene: "ramp4"; product: "ribosome associated membrane protein RAMP4"; Rattus norvegicus mRNA for ribosome associated membrane protein RAMP4, N = 1, Score = 331, P = 6.2e-30

TREMBL:AF100470_1 gene: "RAMP4"; product: "ribosome attached membrane protein 4"; Rattus norvegicus ribosome attached membrane protein 4 (RAMP4) mRNA, complete cds., N = 1, Score = 331, P = 6.2e-30

>TREMBLNEW:RNO238236_1 gene: "ramp4"; product: "ribosome associated membrane protein RAMP4"; Rattus norvegicus mRNA for ribosome associated membrane protein RAMP4
Length = 75

HSPs:

Score = 331 (49.7 bits), Expect = 6.2e-30, P = 6.2e-30
Identities = 66/66 (100%), Positives = 66/66 (100%)

Query: 1 MVAKQIRIRMANEKHSKNITQ RGNVAKTSRNAPEEKASVGPWLLALFIFVVC GSAIFQIIQ 60
MVAKQIRIRMANEKHSKNITQ RGNVAKTSRNAPEEKASVGPWLLALFIFVVC GSAIFQIIQ
Sbjct: 10 MVAKQIRIRMANEKHSKNITQ RGNVAKTSRNAPEEKASVGPWLLALFIFVVC GSAIFQIIQ 69

Query: 61 SIRMGM 66
SIRMGM
Sbjct: 70 SIRMGM 75

No Pedant data available

DKFZphtes3_21n23

group: testes derived

DKFZphtes3_15j18 encodes a novel 148 amino acid protein with strong similarity to rat 7acomp protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

strong similarity to rat 7acomp protein

on genomic level encoded by AF107885

Sequenced by LMU

Locus: /map="14q24.3"

Insert length: 3122 bp

Poly A stretch at pos. 3070, polyadenylation signal at pos. 3045

```
1  GGAAACCTC GTGGGCTCAG CCCGGGAGAA AGGGCCAGGG AAGTTGGGTG
51  GTTCTGTGCT TGGTCTGTCA ATGGAGGAGA TCAAAGTTTT ACGAAGGGTG
101  AAGGAGGAGA ATGATCGGCG AGGTGGATTT ATTCGCATAT TTCCTACATC
151  TGAGACATGG GAAATATATG GGTCTTACCT CGAGCATAAG ACCTCAATGA
201  ACTATATGCT GGCAACACGC CTCTTCCAGG ACAGGGGAAA CCCAAGAAGA
251  AGCTTATTTGA CAGGAAGAAC ACGAATGACT GCTGATGGAG CGCCAGAATT
301  GAAGATAGAG AGTCTGAATT CAAAGGCCAA GCTGCATGCT GCACTTTACG
351  AGAGGAAGCT CCTGTCTCTG GAGGTGCGAA AACGTAGACG ACGGAGTAGC
401  AGATTGAGGG CAATGAGGCC AAAATACCCA GTGATTACCC AACCAGCTGA
451  AATGAATGTT AAAACTGAGA CAGAGAGTGA AGAGGAGGAA GAAGTCGCAT
501  TAGATAATGA AGATGAAGAA CAGGAGGCTT CCCAGGAGGA GTCTGCAGGA
551  TTTCTTAGAG AAAATCAAGC CAAATATACA CCCTCATTGA CAGCTTTGGT
601  AGAAAATACA CCCAAAGAAA ATTCCATGAA AGTTCGTGAA TGGAAATAATA
651  AAGGTGGACA CTGCTGCAAA CTTGAGACTC AGGAGCTAGA GCCTAAATTT
701  AACCTGATGC AGATTCTTCA AGATAATGGC AATCTTAGCA AAATGCAGGC
751  CCGAATAGCA TTCTCTGCCT ATCTCCAGCA TGTTCAAATT CGCCTGATGA
801  AAGACAGTGG CGGTGAGACG TTCAGTGCCA GTTGGGCTGC CAAAGAGGAT
851  GAACAGATGG AGCTGGTTGT TCGTTTCCTC AAGCGAGCAT CAAATAACCT
901  CCAGCATTCA CTGAGGATGG TATTACCCAG TCGACGATTG GCACTTCTGG
951  AACGCAGAAG AATCCTGGCC CACCAGCTGG GTGACTTTAT CATTGTATAC
1001  AACAAAGGAAA CAGAACAAAT GGCTGAAAAG AAATCAAAGA AGAAAGTTGA
1051  GGAAGAAGAG GAAGATGGGG TGAATATGGA AAACCTTCAG GAGTTTCATCA
1101  GACAAGCAAG TGAGGCTGAA CTGGAGGAGG TGTTGACTTT TTATACCCAA
1151  AAGAACAATG CTGCTAGTGT CTTCCTGGGG ACTCACTCTA AAATTTCTAA
1201  GAACAACAAC AATTATTCTG ATAGTGGGGC AAAAGGTGAT CACCCTGAGA
1251  CTATAATGGA AGAAGTGAAA ATAAAGCCAC CTAACAGCA ACAGACGACA
1301  GAAATTCATT CTGATAAATT ATCTCGATTT ACCACTTCAG CAGAAAAAGA
1351  GGCAAAATTA GTTTATAGCA ATTCCTCCTC TGGTCCCTACT GCTACTCTGC
1401  AGAAATTTCC CAACCCCAT TTGTCTCTG TTACAACCTC TGACCTCTCT
1451  CCAGGGCCTT GCCACCATTC TTCTTTATCT CAAATTCCTT CAGCTATCCC
1501  CAGCATGCCCT CACCAGCCAA CAATTTTACT GAACACAGTC TCTGCCAGTG
1551  CTTCTCCCTG CCTACATCCC GGGGCACAGA ACATCCCAAG CCCTACTGGC
1601  CTGCCACGCT GTCGATCAGG AAGTCACACC ATTGGTCCCT TTTCTTCCTT
1651  CCAAAGTGCT GCACACATCT ATAGCCAGAA ACTGTCTCGT CCCTCTTCAG
1701  CAAAGGCAGG ATCGTGCTAT CTAACAAGC ATCATTGAGG AATAGCCAAA
1751  ACACAAAAAG AGGGAGAAGA TGCTTCTTTA TATAGCAAAC GGTACAACCA
1801  AAGTATGGTT ACAGCTGAAC TTCAGCGGCT AGCTGAGAAG CAGGCAGCGA
1851  GACAGTATTC TCCATCCAGC CACATCAACC TCCTCACCCA ACAGGTAACA
1901  AACCTGAATT TGGCAACTGG CATCATAAAC AGAAGCAGTG CTTCAGCTCC
1951  CCCAACCCCTC CGACCCATCA TCAGTCCTAG TGGCCCGACA TGGTCTACAC
2001  AGTCAGACCC CCAAGCTCCC GAGAATCACT CCAGCTCTCC TGGAAGCAGG
2051  AGCCTGCAGA CAGGGGGATT TGCCTGGGAA GGAGAAGTAG AAAACAACGT
2101  GTACAGCCAG GCTACAGGGG TGGTCCCCCA GCACAAGTAT CACCCACAG
2151  CAGGCAGCTA TCAGCTTCAA TTTGCCCTGC AGCAACTTGA ACAACAAAAA
2201  CTTCACTGCC GGCAGCTCCT GGACCAGAGT CGAGCCCGGC ACCAGGCAAT
2251  CTTTGGCAGC CAGACACTAC CTAACCTCAA TTTATGGACA ATGAATAATG
2301  GTGCAGGTTG TAGAATTTCC AGTGCCACAG CTAGTGGCCA GAAGCCAACC
2351  ACTCTGCCAC AAAAAGTGGT ACCACCTCCA AGTTCTTGCG CCTCCCTGGT
2401  TCCCAAAACC CCACCCAACC ACGAACAAGT GCTCAGAAGG GCAACATCCC
2451  AGAAAGCTTC CAATACCCGC TTCAGATCCT CCTTTCAAAA CTATTTGTGG
2501  TATTTCTTCC AAGCAGTCAG CTGAAGTCTG GACGACAGCC TACAACAAC
2551  TACATGCATC TGAAGTGTCT CTTGTAAATG AGCTTTTTTC AGAGCCAGAA
2601  TCATACTCTC CAGGAAATAT GGAGAAAGAA ACCTGAGGAG ATTGAAGTTT
2651  GCCAGGCACA AGGGCAAAAC TCAGACTGAA TGAATTTGAA AGGGTGGGGC
2701  CAAAGATGTT GTAACTGGG AGACTTCTCT GAAGAAAGAA AACTGTTTAA
```

```

2751 GAAACACAGA CTGAAGTCA GTACTTTTCC TTAATAGCT GAGATGACCT
2801 TCTTTACCCCT GGGCTTAGGT GATTCTCATC AGGGTGACCT GAGTGGGAAGT
2851 TGGTGGTAAC GACTGTTCTG TGTCAGCACC CAGGACAGTG GTGTCTGTTA
2901 AGGCTGCCAG GGATTAGCAG GGAGGAAAGC CATCAGGACT GGGTAGCCTG
2951 GTAGCACCAA ATCCCAATTA ATGTTACCTG AACATGTGGT GAGGTCAGCC
3001 GTATGATGAA AGATGTTTAA GAGATTAATG TCAGAAGAAT ATGAAAATAA
3051 ACACCGCGCTT AAAAAATGTT AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3101 AAAAAAAAAA AAAAAAAAAA AA

```

BLAST Results

Entry AF107885 from database EMBL:
Homo sapiens chromosome 14q24.3 clone BAC270M14 transforming growth factor-beta 3 (TGF-beta 3) gene, complete cds; and unknown genes.
Score = 3042, P = 3.0e-219, identities = 610/612
5 exons matching 1893-3070

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 71 bp to 2521 bp; peptide length: 817
Category: strong similarity to known protein

```

1 MEEIKVLRV KEENDRRGGF IRIFPTSETW EIYGSYLEHK TSMNYMLATR
51 LFQDRGNPRR SLTGRTTRMT ADGAPELKIE SLNSKAKLHA ALYERKLLSL
101 EVRKRRRRSS RLRMRPKYP VITQPAEMNV KTETSEEEEE EVALDNEDEE
151 QEASQESAG FLRENQAKYT PSLTALVENT PKENSMKVRE WNNKGGHCCK
201 LETQLEPKF NMQILQDNG NLSKMQARIA FSAYLQHVQI RLMKDSGGQT
251 FSASWAAKED EQMELVVRF LKRASNNLQHS LRMVLPSSRL ALLERRRILA
301 HQLGDFIIVY NKETEQMAEK KSKKKVEEEE EDGVNMENFQ EFIRQASEAE
351 LEEVLTFTYQ KNKSASVFLG THSKISKNNN NYSDSGAKGD HPETIMEEVK
401 IKPPKQQQT EIHSKLSRF TTSAEKEAKL VYSNSSSGPT ATLQKIPNTH
451 LSSVTTSDL SGPCHHSSLS QIPSAIPSM P HQTILLNTV SASASPCLHP
501 GAQNIPSPPT LPRCRSGSHT IGPFSFQSA AHYISQKLSR PSSAKAGSCY
551 LNKHHSGIAK TQKEGEDASL YSKRYNOSMV TAELORLAEK QAARQYSPSS
601 HINLLTQQVT NLNLATGIIN RSSASAPPTL RPIISPSGPT WSTQSDPQAP
651 ENHSSSPGSR SLQTGGFAWE GEVENNVYSQ ATGVVPQHKY HPTAGSYQLQ
701 FALQLEQKQ LQSRQLLDQS RARHQAIFGS QTLPSNLWT MNNGAGCRIS
751 SATASGQKPT TLPQKVPPP SSCASLVKPK PPNHEQVLRR ATSQKASNTR
801 FRSSFQNYLW YFFQAVS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_21n23, frame 2

TREMBL:AF064856_1 product: "7acomp protein"; Rattus sp. 7acomp protein mRNA, complete cds., N = 1, Score = 1845, P = 2.2e-190

TREMBL:AF107885_3 product: "unknown"; Homo sapiens chromosome 14q24.3 clone BAC270M14 transforming growth factor-beta 3 (TGF-beta 3) gene, complete cds; and unknown genes., N = 1, Score = 443, P = 5.3e-41

TREMBL:AF107885_4 product: "unknown"; Homo sapiens chromosome 14q24.3 clone BAC270M14 transforming growth factor-beta 3 (TGF-beta 3) gene, complete cds; and unknown genes., N = 1, Score = 265, P = 8.2e-22

>TREMBL:AF064856_1 product: "7acomp protein"; Rattus sp. 7acomp protein mRNA, complete cds.
Length = 436

HSPs:

Score = 1845 (276.8 bits), Expect = 2.2e-190, P = 2.2e-190
Identities = 369/435 (84%), Positives = 395/435 (90%)

Query: 115 MRPKYPVITQPAEMNVKTETESEEEEVALDNEDEEQEASQESAGFLRENQAKYTPSLT 174
MRPKYPVIT PAEMN+KTETESEEEEV LDNEDEEQEASQESAG L ENQAKYTPSLT
Sbjct: 1 MRPKYPVITLPAEMNIKTETESEEEEVLDNEDEEQEASQESAGSLAENQAKYTPSLT 60

Query: 175 ALVENTPKENSMKVREWNKGGHCKKLETQELEPKFNLQMILQDNGNLSKMQARIAFSAY 234
+VEN+P+EN+MKV EW NKG CCK+ETQE E KFNLMQILQDNGNLSK+QAR+AFSAY
Sbjct: 61 VIVENSRENAMKVAEWTNKGESCKIETQEPESKFNLMQILQDNGNLSKVQARLAFSAY 120

Query: 235 LQHVQIRLMKDSGGQTFASWAAKEDQEMELVVRFLKRASNQLQSLRMVLPSSRLALLE 294
LQHVQ+RL KDSGGQT S SWAAKEDQEMELVVRFLKRAS+NLQSLRMVLPSSRLALLE
Sbjct: 121 LQHVQVRLTKDSGGQTLSPSWAAKEDQEMELVVRFLKRASNQLQSLRMVLPSSRLALLE 180

Query: 295 RRRILAHQLGDFIIVYNKETEQAEEKSKKKVEEEEEDGVNMENFQEFIRQASEAELEEV 354
RRRILAHQLGDFI+VYNKETEQAEEKSKKK+EEEEEDGVN E+FQEFIRQASEAELEEV
Sbjct: 181 RRRILAHQLGDFIIVYNKETEQAEEKSKKKLEEEEEDGVNAESFQEFIRQASEAELEEV 240

Query: 355 LTFYQTQKNKSASVFLGTHSKISKNNNNYSDSGAKGDHPETIMEEVKIKPPKQQQTTEIHS 414
LTFYQTQKNKSASVFLGTHSK SKN+++YSDSGAKGDHPETI +EVKIK PKQQQ TEIHS
Sbjct: 241 LTFYQTQKNKSASVFLGTHSKSSKNSSSYSDSGAKGDHPETI-QEVKIKQPKQQQATEIHS 299

Query: 415 DKLSRFTTSAEKEAKLVYSNSSS--GPTATL-QKIPNTHLSSV-TTSDLSPGPCHHSSLS 470
DKLSRFTTSA KEAKLVY+N SS GP A L Q+P+THLSS+ TTS LS GP HHSSLS
Sbjct: 300 DKLSRFTTSAEKEAKLVYTNCSSFSGPAAVLLQRLPSTHLSSIITSTLSSGPGHHSSLS 359

Query: 471 QIPSAIPSMHPQPTILLNTVSASASPCLPQAQNPSTGLPCRSGSHTIGPFSSSFQSA 530
QI AIPSMHPQ +LLN V SASP +HPG N+ SP GLPCRSGS+TIGPFSSSFQSA
Sbjct: 360 QISPAIPSMHPQSALLNPVPDSASPPVHPGTPNV-SPAGLPCRSGSYTIGPFSSSFQSA 418

Query: 531 AHIYSQKLSRPSSAKAG 547
AHIYSQKLSRPSSAKAG
Sbjct: 419 AHIYSQKLSRPSSAKAG 435

Pedant information for DKFZphtes3_2ln23, frame 2

Report for DKFZphtes3_2ln23.2

[LENGTH] 817
[MW] 91522.09
[pI] 9.32
[HOMOL] TREMBL:AF064856_1 product: "7a comp protein"; Rattus sp. 7a comp protein mRNA,
complete cds. 1e-166
[PROSITE] MYRISTYL 6
[PROSITE] CAMP_PHOSPHO_SITE 4
[PROSITE] CK2_PHOSPHO_SITE 12
[PROSITE] TYR_PHOSPHO_SITE 1
[PROSITE] PKC_PHOSPHO_SITE 15
[PROSITE] ASN_GLYCOSYLATION 7
[KW] Alpha Beta
[KW] LOW_COMPLEXITY 13.83 %

SEQ MEEIKVLRVKEENDRRGGFIRIFPTSETWEIYGSYLEHKTSMNMYLATRLFQDRGNPRR
SEG
PRD cccchhhhhhhhhccccceeeccccceeeccccceeeccccchhhhhhhhhhhcccccc

SEQ SLLTGRTRMTADGAPELKIESLNSKAKLHAALYERKLLSLEVRKRRRRSSRLRAMRPKYP
SEGxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.....
PRD cccccceeeccccceeeehhhcccc

SEQ VITQPAEMNVKTETESEEEEVALDNEDEEQEASQESAGFLRENQAKYTPSLTALVENT
SEGxx.....
PRD ceeecchhhhhhhhhhhhhhhhhhhhhccchhhhhhhhhhhhhhhhhhhhhccccceeecccc

SEQ PKENSMKVREWNKGGHCKKLETQELEPKFNLQMILQDNGNLSKMQARIAFSAYLQHVQI
SEG
PRD cccccceeeccccccccchhhhhhhccchhhhhhhccchhhhhhhhhhhhhhhhhhhhh

SEQ RLMKDSGGQTFASWAAKEDQEMELVVRFLKRASNQLQSLRMVLPSSRLALLERRRILA
SEGxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.....
PRD hhhccccceeehh

SEQ HQLGDFIIVYNKETEQAEEKSKKKVEEEEEDGVNMENFQEFIRQASEAELEEVLTFTYQ
SEGxxxxxxxxxxxxxxxx.....
PRD hhccccceeehhhhhhhhhhhhhhhhhhhhhhccchhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ KNKSASVFLGTHSKISKNNNNYSDSGAKGDHPETIMEEVKIKPPKQQQTTEIHSKLSRF
SEG
PRD cccccceeeccccccccccccccccccccccccchhhhhhhccccccccceeecccccccc

```

SEQ      TTSAEKEAKLVYSNSSSGPTATLQKIPNTHLSSVTTSDLSPGPGCHHSSLSQIPSAIPSPMP
SEG      .....
PRD      hhhhhhhheeeccccccccceeecccccccccccccccccccccccccccccccccccccc

SEQ      HQPTILNTVSASASPLHPGAQNIPSTGLPCRSGSHTIGPFSSFQSAAHYYSQKLSR
SEG      .....
PRD      cccceeeccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhccc

SEQ      PSSAKAGSCYLNKHHSGIAKTQKEGEDASLYSKRYNQSMVTAEQLRLAEKQAAQYSPSS
SEG      .....
PRD      cccccceeeccccccccccccccccccccceeeccchhhhhhhhhhhhhhhhhhhhhccc

SEQ      HINLLTQQVTNLNLATGIINRSSASAPPTLRPIISPSGPTWTSQSDPQAPENHSSSPGSR
SEG      ..xxxxxxxxxxxx.....xxxxxxxxxxxxxxxxxxxxxxxxxxxx...
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ      SLQTGGFAWEGEVENNVYSQATGVVPQHKYHPTAGSYQLQFALQQLEQQKLQSRQLLDQS
SEG      .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx...
PRD      cccccceeeccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhh

SEQ      RARHQAIFGSQTLPNSNLWTMNGAGCRISSATASGQKPTTLPQKVVPFPSSCASLVPKP
SEG      .....
PRD      hhhhhhhccccccccceeeccccceeeccccccccceeeccccceeecccc

SEQ      PPNHEQVLRRTATSQKASNTRFRSSFQNYLWYFFQAVS
SEG      .....
PRD      cccchhhhhhhhhhhccccccccccccceeeeeecccc

```

Prosites for DKFZphtes3_2ln23.2

PS00001	221->225	ASN_GLYCOSYLATION	PDOC00001
PS00001	362->366	ASN_GLYCOSYLATION	PDOC00001
PS00001	381->385	ASN_GLYCOSYLATION	PDOC00001
PS00001	434->438	ASN_GLYCOSYLATION	PDOC00001
PS00001	576->580	ASN_GLYCOSYLATION	PDOC00001
PS00001	620->624	ASN_GLYCOSYLATION	PDOC00001
PS00001	652->656	ASN_GLYCOSYLATION	PDOC00001
PS00004	106->110	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	107->111	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	271->275	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	789->793	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	64->67	PKC_PHOSPHO_SITE	PDOC00005
PS00005	109->112	PKC_PHOSPHO_SITE	PDOC00005
PS00005	180->183	PKC_PHOSPHO_SITE	PDOC00005
PS00005	185->188	PKC_PHOSPHO_SITE	PDOC00005
PS00005	280->283	PKC_PHOSPHO_SITE	PDOC00005
PS00005	287->290	PKC_PHOSPHO_SITE	PDOC00005
PS00005	322->325	PKC_PHOSPHO_SITE	PDOC00005
PS00005	359->362	PKC_PHOSPHO_SITE	PDOC00005
PS00005	414->417	PKC_PHOSPHO_SITE	PDOC00005
PS00005	535->538	PKC_PHOSPHO_SITE	PDOC00005
PS00005	543->546	PKC_PHOSPHO_SITE	PDOC00005
PS00005	561->564	PKC_PHOSPHO_SITE	PDOC00005
PS00005	572->575	PKC_PHOSPHO_SITE	PDOC00005
PS00005	629->632	PKC_PHOSPHO_SITE	PDOC00005
PS00005	793->796	PKC_PHOSPHO_SITE	PDOC00005
PS00006	35->39	CK2_PHOSPHO_SITE	PDOC00006
PS00006	132->136	CK2_PHOSPHO_SITE	PDOC00006
PS00006	134->138	CK2_PHOSPHO_SITE	PDOC00006
PS00006	136->140	CK2_PHOSPHO_SITE	PDOC00006
PS00006	154->158	CK2_PHOSPHO_SITE	PDOC00006
PS00006	180->184	CK2_PHOSPHO_SITE	PDOC00006
PS00006	347->351	CK2_PHOSPHO_SITE	PDOC00006
PS00006	394->398	CK2_PHOSPHO_SITE	PDOC00006
PS00006	422->426	CK2_PHOSPHO_SITE	PDOC00006
PS00006	455->459	CK2_PHOSPHO_SITE	PDOC00006
PS00006	561->565	CK2_PHOSPHO_SITE	PDOC00006
PS00006	643->647	CK2_PHOSPHO_SITE	PDOC00006
PS00007	563->572	TYR_PHOSPHO_SITE	PDOC00007
PS00008	195->201	MYRISTYL	PDOC00008
PS00008	248->254	MYRISTYL	PDOC00008
PS00008	510->516	MYRISTYL	PDOC00008
PS00008	557->563	MYRISTYL	PDOC00008
PS00008	746->752	MYRISTYL	PDOC00008
PS00008	756->762	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_2ln23.2)

DKFZphtes3_22c23

group: testes derived

DKFZphtes3_22c23 encodes a novel 223 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, 3 EST hits (two from a testis library)

Sequenced by LMU

Locus: /map="9q34"

Insert length: 1113 bp

Poly A stretch at pos. 1073, polyadenylation signal at pos. 1055

```

1 GGTGGGCAAA GGCATCTTCC TCTGGGAAGG ACTGGCACAA GCACTTGGTC
51 CCTGGGTTGT GTGCCTGGGA GGCCGGGATC AGGGCTGGCC CTCTTTCTCC
101 CTGGCAAAGC AAAACCTCCC TTTTACTACT ATCAAGGGA AGTAACTTGA
151 AGGTGCCTGT GGCAGGCAGC ACCTTGAGCC AACAGGAACC ATTGACATGC
201 GAGGCCAGG GCAGGCAGAC TGTGCAGTGG CCATTGGCGC GCCCCTCGGG
251 GAGGTGGTGA CCCTCCGCGT CCTTGAGAGT TCTCTCAACT GCAGTGGCGG
301 GGACATGTTG CTGCTTTGGG GCCGGCTCAC CTGGAGGAAG ATGTGCAGGA
351 AGCTGTTGGA CATGACTTTC AGCTCCAAGA CCAACACGCT GGTGGTGAGG
401 CAGCGCTGCG GGCGGCCAGG AGGTGGGGTG CTGCTGCGGT ATGGGAGCCA
451 GCTTGCTCCT GAAACCTTCT ACAGAGAATG TGACATGCAG CTCTTTGGGC
501 CCTGGGGTGA AATCGTGAGC CCCTCGCTGA GTCCAGCCAC GAGTAATGCA
551 GGGGGCTGCC GGCTCTTCAT TAATGTGGCT CCGCACGCAC GGATTGCCAT
601 CCATGCCCTG GCCACCAACA TGGGCGCTGG GACCGAGGGA GCCAATGCCA
651 GGTACATCTT GATCCGGGAC ACCACAGCT TGAGGACCAC AGCGTTCCAT
701 GGGCAGCAGG TGCTCTACTG GGAGTCAGAG AGCAGCCAGG CTGAGATGGA
751 GTTCAGCGAG GGCTTCTGTA AGGCTCAGGC CAGCCTGCGG GGCCAGTACT
801 GGACCCTCCA ATCATGGGTA CCGGAGATGC AGGACCCTCA GTCCTGGAAG
851 GGAAAGGAAG GAACCTGAGG GTCATTGAAC ATTTGTCCG TGTCTGGCCA
901 GCCCTGGAGG GTTGACCCCT GGTCTCAGTG CTTTCCAATT CGAACTTTT
951 CCAATCTTAG GTATCTACTT TAGAGTCTTC TCCAATGTCC AAAAGGCTAG
1001 GGGGTGGAG GTGGGGACTC TGGAAAAGCA GCCCCATT CTCTGGGTAC
1051 CAATAAATAA AACATGCAGG CTGAAAAAAA AAAAAAAAAA AAAAAAAAAA
1101 AAAAAAAAAA AAA

```

BLAST Results

Entry HSAC1644 from database EMBL:
 Genomic sequence from Human 9q34, complete sequence.
 Score = 2072, P = 8.8e-225, identities = 422/430
 5 exons Bp 41969-38232

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 197 bp to 865 bp; peptide length: 223
 Category: putative protein

```

1 MRPGQADCA VAIGRPLGEV VTLRVLESSL NCSAGDMLLL WGRLTWRKMC
51 RKLLDMTFSS KNTLVVRQR CGRPGGGVLL RYGSQAPET FYRECDMLF
101 GPWGEIVSPS LSPATSNAGG CRLFINVAPH ARIAIHALAT NMAGTEGAN
151 ASYLIRDTLH SLRTTAFHGQ QVLYWESESS QAEMEFSEGF LKAQASLRGQ
201 YWTLQSWVPE MQDPQSWKKG EGT

```

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3 22c23, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphtes3 22c23, frame 2

Report for DKFZphtes3 22c23.2

[illegible]

Prosites for DKFZphtes3 22c23.2

PS000001	31->35	ASN_GLYCOSYLATION	PDOC000001
PS000001	150->154	ASN_GLYCOSYLATION	PDOC000001
PS000005	22->25	PKC_PHOSPHO_SITE	PDOC000005
PS000005	45->48	PKC_PHOSPHO_SITE	PDOC000005
PS000005	59->62	PKC_PHOSPHO_SITE	PDOC000005
PS000005	161->164	PKC_PHOSPHO_SITE	PDOC000005
PS000005	196->199	PKC_PHOSPHO_SITE	PDOC000005
PS000005	216->219	PKC_PHOSPHO_SITE	PDOC000005
PS000006	33->37	CK2_PHOSPHO_SITE	PDOC000006
PS000006	180->184	CK2_PHOSPHO_SITE	PDOC000006
PS000008	5->11	MYRISTYL	PDOC000008
PS000008	145->151	MYRISTYL	PDOC000008
PS000008	148->154	MYRISTYL	PDOC000008
PS000008	199->205	MYRISTYL	PDOC000008

(No Pfam data available for DKFZphtes3 22c23.2)

DKFZphtes3_22g2

group: nucleic acid management

DKFZphtes3_22g2 encodes a novel 1230 amino acid protein with nearly identical to rat TIP120.

TATA-binding protein TBP is a central component for transcriptional regulation and is a target for various transcription regulators. TBP-interacting protein 120 (TIP120) is a protein interacting with the TATA-binding protein (TBP). The novel protein is the human ortholog of rat TIP120. The novel TBP-binding protein is considered to participate in transcription regulation through the interaction with TBP.

The new protein can find application in modulation of gene transcription.

KIAA0829, complete cds, nearly identical to rat TIP120

complete cDNA, complete cds, EST hits,

Sequenced by LMU

Locus: /map="387.3 cR from top of Chr12 linkage group"

Insert length: 5387 bp

Poly A stretch at pos. 5352, polyadenylation signal at pos. 5335

```
1 GGGAGCGAGT GCGGAGCGAG TGGGAGCGAG ACGGCCCTGA GTGGAAGTGT
51 CTGGCTCCCC GTAGAGGCCG TTCTGTACGC CCCGCCGCCG ATGAGCTCGT
101 TCTCACGCCG ACAGCGCCGT CGTTAGGCTG GCTCTGTAGC CTCGGCTTAC
151 CCCGGGACAG GCCCAGCCCT CGCCAGGGAG GGGGCAGCCC GTCGAGGCGC
201 CTCCCTAGTC AGCGTCGGCG TCGCGCTCGG ACCCTGGAAG CGGGAGCCGC
251 CGCGAGCGAG AGGAGGAGCT CCAGTGGCGG CGGCGGCGGC GGCAGCGGCA
301 GCGGGCAGCA GCTCCAGCAG CGCCAGCAGG CGGGATCGAG GCCGTCAACA
351 TGGCGAGCGC CTCGTACCAC ATTTCCAATT TGCTGGAATA AATGACATCC
401 AGCGACAAGG ACTTTAGGTT TATGGCTACA AATGATTGTA TGACGGAACT
451 GCAGAAAGAT TCCATCAAGT TGGATGATGA TAGTGAAAGG AAAGTAGTGA
501 AAATGATTTT GAAGTTATTG GAAGATAAAA ATGGAGAGGT ACAGAAATTA
551 GCTGTCAAAAT GTCTTGGTCC TTTAGTGAGT AAAGTGAAAG AATACCAAGT
601 AGAGACAATT GTAGATACCC TCTGCACTAA CATGCTTTCT GATAAAGAAC
651 AACTTCGAGA CATTTCAAGT ATTGGTCTTA AAACAGTAAT TGGAGAACTT
701 CCTCCAGCTT CCAGTGGCTC TGCATTAGCT GCTAATGTAT GTAAAAAGAT
751 TACTGGACGT CTTACAAGTG CAATAGCAAA ACAGGAAGAT GTCTCTGTTC
801 AGCTAGAAGC CTTGGATATT ATGGCTGATA TGTGAGCAGG GCAAGGAGGA
851 CTTCTTGTTA ATTTCCATCC TTCAATTCTG ACCTGTCTAC TTCCCCAGTT
901 GACCAAGCCG AGACTTGCAG TGAGGAAAAG AACCATTATC GCTCTTGGCC
951 ATCTGGTTAG GAGCTGTGGA AATATAGTTT TTGTAGATCT TATTGAACAT
1001 CTGTTGTCTG AGTTGTCCAA AAATGATTCT ATGTCAACAA CAAGAACCTA
1051 CATACAATGT ATTGTGCTA TTAGTAGGCA AGCTGGTCAT AGAATAGGTG
1101 AATACCTTGA GAAGATAATT CCTTTGGTGG TAAAAATTTG CAATGTAGAT
1151 GATGATGAAT TAAGAGAGTA CTGTATTCAA GCCTTTGAAT CATTGTGAAG
1201 AAGATGTCTT AAGGAAGTAT ATCCTCATGT TTCTACCATT ATAAATATTT
1251 GTCTTAAATA TCTTACCTAT GATCCAAATT ATAATTACGA TGATGAAGAT
1301 GAAGATGAAA ATGCAATGGA TGCTGATGGT GGTGATGATG ATGATCAAGG
1351 GAGTGATGAT GAATACAGTG ATGATGATGA CATGAGTTGG AAAGTGAGAC
1401 GTGCAGCTGC GAAAGTCTTG GATGCTGTAG TTAGCACAAG GCATGAAATG
1451 CTTCCAGAAAT TCTACAAGAC CGTCTCTCCT GCACTAATAT CCAGATTTAA
1501 AGAGCGTGAA GAGAAATGTA AGGCAGATGT TTTTCACGCA TACCTTTCTC
1551 TTTTGAAGCA AACTCGTCCT GTACAAAGTT GGCTATGTGA CCCTGATGCA
1601 ATGGAGCAGG GAGAAACACC TTTAACAATG CTTCAGAGTC AGGTTCCCAA
1651 CATTGTTAAA GCTCTTCACA AACAGATGAA AGAAAAAAGT GTGAAGACCC
1701 GACAGTGTG TTTTAAACATG TTAACGTAGC TGGTAAATGT ATTACCTGGG
1751 GCCCTAATC AACACATTCC TGTACTTGTA CCAGGAATCA TTTTCTCACT
1801 GAATGATAAA TCAAGCTCAT CGAATTTGAA GATCGATGCT TTGTCATGTC
1851 TATACGTAAT CCTCTGTAAC CATTCTCCTC AAGTCTTCCA TCCTCACGTT
1901 CAGGCTTTGG TTCCTCCAGT GGTGGCTTGT GTTGGAGACC CATTTTACAA
1951 AATTACATCT GAAGCACTTC TTGTTACTCA ACAGCTTGTC AAAGTAATTC
2001 GTCCTTTTGA TCAGCCTTCC TCGTTTGATG CAACTCCTTA TATCAAAGAT
2051 CTATTTACCT GTACCATTAAG GAGATTAAAA GCAGCTGACA TTGATCAGGA
2101 AGTCAAGGAA AGGGCTATTT CCTGTATGGG ACAAATTATT TGCAACCTTG
2151 GAGACAATTT GGGTTCTGAC TTGCCTAATA CACTTCAGAT TTTCTTGGAG
2201 AGACTAAAGA ATGAAATTAC CAGGTTAACT ACAGTAAAGG CATTGACACT
2251 GATTGCTGGG TCACCTTTGA AGATAGATTG GAGGCCTGTT CTGGGAGAAG
2301 GGGTTCTTAT CCTTGCTTCA TTTCTTAGAA AAAACCAGAG AGCTTTGAAA
2351 CTGGGTACTC TTTCTGCCCT TGATATTCTA ATAAAAAAGT ATAGTGACAG
2401 CTTGACAGCT GCCATGATTG ATGCAGTTCT AGATGAGCTC CCACCTCTTA
2451 TCAGCGAAAG TGATATGCAT GTTTCACAAA TGCCCATCAG TTTTCTTACC
2501 ACTTTGGCAA AAGTATATCC CTCCTCCCTT TCAAAGATAA GTGGATCCAT
2551 TCTCAATGAA CTTATTGGAC TTGTGAGATC ACCCTTATTG CAGGGGGGAG
```

```

2601 CTCTTAGTGC CATGCTAGAC TTTTCCAAG CTCTGGTTGT CACTGGAACA
2651 AATAATTTAG GATACATGGA TTTGTTGCGC ATGCTGACTG GTCCAGTTTA
2701 CTCTCAGAGC ACAGCTCTTA CTCATAAGCA GTCTTATTAT TCATTGGCCA
2751 AATGTGTAGC TGCCCTTACT CGAGCATGCC CTAAGAGGGG ACCAGCTGTA
2801 GTAGGTCAGT TTATTCAAGA TGTCAAGAAC TCAAGGTCTA CAGATTCCAT
2851 TCGTCTCTTA GCTCTACTTT CTCTTGGAGA AGTTGGGCAT CATATTGACT
2901 TAAGTGGACA GTTGGAACTA AAATCTGTAA TACTAGAAGC TTTCTCATCT
2951 CCTAGTGAAG AAGTCAAATC AGCTGCATCC TATGCATTAG GCAGCATTAG
3001 TGTGGGCAAC CTTCCTGAAT ATCTGCCGTT TGTCCCTGAA GAAATAACTA
3051 GTCAACCCAA AAGGCAGTAT CTTTACTTTC ATTCTTGAA GGAAATTATT
3101 AGCTCTGCAT CAGTGGTGGG CCTTAAACCA TATGTTGAAA ACATCTGGGC
3151 CTTATTACTA AAGCACTGTG AGTGTGCAGA GGAAGGAACC AGAAATGTTG
3201 TTGCTGAATG TCTAGGAAAA CTCACTCTAA TTGATCCAGA AACTCTCCTT
3251 CCACGGCTTA AGGGGTACTT GATATCAGGC TCATCATATG CCCGAAGCTC
3301 AGTGGTTAGC GCTGTGAAAT TTACAATTTT TGACCATCCA CAACCTATTG
3351 ATCCACTGTT AAAGAACTGC ATAGGTGATT TCCTAAAAAC TTTGGAAGAC
3401 CCAGATTGGA ATGTGAGAAG AGTAGCCTTG GTCACATTTA ATTCAGCAGC
3451 ACATAACAAG CCATCATTAA TAAGGGATCT ATTGGATACT GTTCTTCCAC
3501 ATCTTTTACA TGAAACAAAA GTTAGAAAGG AGCTTATAAG AGAGGTAGAA
3551 ATGGGTCCAT TTAACATAC GGTGTATGAT GGTCTGGATA TTAGAAAGGC
3601 AGCATTTGAG TGTATGTACA CACTTCTAGA CAGTTGTCTT GATAGACTTG
3651 ATATCTTTGA ATTTCTAAAT CATGTTGAAG ATGTTTGAA GGACCATTAT
3701 GATATTAAAG TGCTGACATT TTTAATGTTG GTGAGACTGT CTACCCTTTG
3751 TCCAAGTGCA GTACTGCAGA GGTGGACCG ACTTGTGAG GCATTACGTG
3801 CAACATGTAC AACTAAGGTA AAGGCAAACT CAGTAAAGCA GGAGTTTGAA
3851 AAACAAGATG AATTAAAGCG ATCTGCCATG AGAGCAGTAG CAGCACTGCT
3901 AACCATTCCA GAAGCAGAGA AGAGTCCACT GATGAGTGAA TTCCAGTCAC
3951 AGATCAGTTC TAACCCTGAG CTGGCGGCTA TCTTTGAAAG TATCCAGAAA
4001 GATTTCATCAT CTAATACTTT GGAATCAATG GACACTAGTT AGATGTTTGT
4051 TCACCATGGG GACCATTAACA TATGACCATA CAATGCACTG AATTGACAGG
4101 TTAATCATAA GACATGGAAA GAGAAGTGTC TAAAGCTTC AAAATGTTCC
4151 ACTTTTTTTT CCTTCATGGA GACTGTTTGT TTGGCTTTCT TCCATTGTTG
4201 TTTTGTAGC ATTTATTTCA GAAATGTGTA TTCCATAAT CCAGAGGTTG
4251 TAAACCACT AGTGTTTTAG TGGTTACAGC AACATTGAA ATGGAAACTA
4301 AAAGTTAGGA TTTTATGGAG TATGGAGATA GGTCCAGTA TCTATTACC
4351 CTGTAATGTT TAGGATTAAA ATGTTAAAT TTTGTGACCA TGAATTCTT
4401 TCTTTTATAA ATTTCTCAT TTAATAATCA AAAATCTGC AAAACAAAAA
4451 CCATGTTTCT TTTCTTGTA TAACTTTGT TTTTCAGCAA CATAAATTGA
4501 TTTTATAGCT GCAGACAAGA ATATCCATAT AAGATTGTTT AACCATTTCA
4551 GAGAGTTTGG CAATTTTAA AAGATAATAA GGTATCATTT TTAAGTATGA
4601 AAATTAACAA TATCCCTGTT GCGCACACTA ATTTTGCATG AGTAAGTTTA
4651 CAAATATGTA TCGCTGTGTA AGCAGCATGT GCAGATTATT CATAATATAG
4701 AAGTTAAAT AAGTATTAGT GCAATTTTCA GATATTATT TTTGCACAGA
4751 AAACACATTA TCTGGAGAGA AAGAAAGGAG AATTTTGTAG ACTTGGGTTT
4801 TCTTAATGCC AGTGTGAATT TGCAGATGTT TTCAGAAAAT CAAGTCACAG
4851 TAACAATTTG CCACTTTTTT CTATTATAAA TCTTCTTACT TAAATTTTGA
4901 ATATTTAGTT TTTCTCAGTT ACCCATTTGT GTGTGTGTGA TTCCACTTAG
4951 AAATCTTAA AACCAGATT TTCTTTCATT CCGTTTGGAT GTCTACATTC
5001 CTTATCAAAG GATATAAATA CTGTGTATGC TTTTGAATTT TATTTTATAG
5051 AAAAATCTGA AGCCAGCTAT CACAGGTTT TTAGCTAATA ATAGATTTT
5101 CTTTATGTTG AGTTAGGTTT TTCCCATCT CCTGTAGAGC GAATTTACAT
5151 ATTGTATTGG GTAAGTGTTC ACTACTTTTC CTGATTAAGG GATCTGTGCT
5201 GGGGAACAAA GCTTTTGCAG TACCTTATAT TGTAGTTAAA ATTTTATTTA
5251 ACATATCCTT CAGTGAGCTC ATTTACACT GTAGCCTCTT CCTTAAATTT
5301 TGTGGTGCCT CTGTAAACAGT AAGAACTAAT TCTGAAATAA AAGACATCTC
5351 CTAAAAAAA AAAAAAAA AAAAAAAA AAAAAA

```

BLAST Results

Entry HS793345 from database EMBL:
human STS WI-12457.
Score = 1985, P = 1.3e-83, identities = 433/460

Medline entries

97127450:
Molecular cloning of a novel 120-kDa TBP-interacting
protein.

Peptide information for frame 2

ORF from 350 bp to 4039 bp; peptide length: 1230

Category: known protein
Classification: Nucleic acid management

```
1 MASASYHISN LLEKMTSSDK DFRFMATNDL MTELQKDSIK LDDDSERKVV
51 KMILKLEDEK NGEVQNLA VK CLGPLVSKVK EYQVETIVDT LCTNMLSDEK
101 QLRDISSIGL KTVIGELPPA SSGSALAANV CKKITGRLTS AIKQEDVSV
151 QLEALDIMAD MLRSQGGLLV NFHPSILTCL LPQLTSPRLA VRKRTIIALG
201 HLVMSCGNIV FVDLIEHLLS ELSKNDSMST TRTYIQCIAA ISRQAGHRIG
251 EYLEKIIPLV VKFCNVDDDE LREYCIQAFE SFVRRCPKEV YPHVSTIINI
301 CLKYLTYPDN YNYDDEDEDE NAMDADGGDD DDQGSDDDEYS DDDMSWKVR
351 RAAAKCLDAV VSTRHEMLPE FYKTVSPALI SRFKEREENV KADV FHAYLS
401 LLKQTRPVQS WLCDDPDAMEQ GETPLTMLQS QVPNIVKALH KQMKESVKT
451 RQCCFNMLTE LVNVLPALT QHIPVLVPGI IFSLNDKSSS SNLKIDALSC
501 LYVILCNHSP QVFHHPVQAL VPPVACVGD PFYKITSEAL LVTQQLVKVI
551 RPLDQPSFSD ATPYIKDLFT CTIKRLKAAD IDQEVKERAI SCMGQICNL
601 GDNLGSDLPN TLQIFLERLK NEITRLTTVK ALTLIAGSPL KIDLRPVLGE
651 GSPILASFLR KNQRAKLGT LSALDILIKN YSDSLTAAMI DAVLDELPLP
701 ISESDMHVSQ MAISFLTTLA KVYPSSLSKI SGSILNELIG LVRSPLLQGG
751 ALSAMLDFFQ ALVVTGTNNL GYMDLLRMLT GPVYSQSTAL THKQSYSSIA
801 KCVAALTRAC PKEGPAVVGQ FIQDVKNRSR TDSIRLLALL SLGEVGHID
851 LSGQLELKSV ILEAFSSPSE EVKSAASYAL GSISVGNLPE YLPFVLQEI
901 SQPKRQYLIL HSLKEIISSA SVVGLKPYVE NIWALLKHC ECAEEGTRNV
951 VAECLGKLT IDPETLLPRL KGYLISGSSY ARSSVVTAVK FTISDHPQPI
1001 DPLLKNCIGD FLKTLEDPLD NVRVALVTF NSAAHNKPSL IRDLLDTVLP
1051 HLYNETKVRK ELIREVEMGP FKHTVDDGLD IRKAAFECMY TLLDSCLDRL
1101 DIFEFLNHVE DGLKDHYDIK MLTFLMLVRL STLCPSAVLQ RLDRLVEPLR
1151 ATCTTKVKAN SVKQEFKQD ELKRSAMRAV AALLTIPEAE KSPLMSEFQS
1201 QISSNPELAA IFESIQRDSS STNLESMDTS
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_22g2, frame 2

TREMBL:AB020636_1 gene: "KIAA0829"; product: "KIAA0829 protein"; Homo sapiens mRNA for KIAA0829 protein, partial cds., N = 1, Score = 5986, P = 0

TREMBL:RND6711_1 gene: "tip120"; product: "TIP120"; Rattus norvegicus mRNA for TIP120, complete cds., N = 1, Score = 6203, P = 0

>TREMBL:RND6711_1 gene: "tip120"; product: "TIP120"; Rattus norvegicus mRNA for TIP120, complete cds.
Length = 1,230

HSPs:

Score = 6203 (930.7 bits), Expect = 0.0e+00, P = 0.0e+00
Identities = 1227/1230 (99%), Positives = 1228/1230 (99%)

```
Query:      1 MASASYHISN LLEKMTSSDK DFRFMATNDL MTELQKDSIK LDDDSERKVV KMILKLEDEK 60
            MASASYHISN LLEKMTSSDK DFRFMATNDL MTELQKDSIK LDDDSERKVV KMILKLEDEK
Sbjct:      1 MASASYHISN LLEKMTSSDK DFRFMATNDL MTELQKDSIK LDDDSERKVV KMILKLEDEK 60

Query:      61 NGEVQNLA VK CLGPLVSKVK EYQVETIVDT LCTNMLSDEK QLRDISSIGL KTVIGELPPA 120
            NGEVQNLA VK CLGPLVSKVK EYQVETIVDT LCTNMLSDEK QLRDISSIGL KTVIGELPPA
Sbjct:      61 NGEVQNLA VK CLGPLVSKVK EYQVETIVDT LCTNMLSDEK QLRDISSIGL KTVIGELPPA 120

Query:      121 SSGSALAANV CKKITGRLTS AIKQEDVSV QLEALDIMAD MLRSQGGLLV NFHPSILTCL 180
            SSGSALAANV CKKITGRLTS AIKQEDVSV QLEALDIMAD MLRSQGGLLV NFHPSILTCL
Sbjct:      121 SSGSALAANV CKKITGRLTS AIKQEDVSV QLEALDIMAD MLRSQGGLLV NFHPSILTCL 180

Query:      181 LPQLTSPRLA VRKRTI IALGHLVMSCGNIV FVDLIEHLL SELSKNDSMST TRTYIQCIAA 240
            LPQLTSPRLA VRKRTI IALGHLVMSCGNIV FVDLIEHLL SELSKNDSMST TRTYIQCIAA
Sbjct:      181 LPQLTSPRLA VRKRTI IALGHLVMSCGNIV FVDLIEHLL SELSKNDSMST TRTYIQCIAA 240

Query:      241 ISRQAGHRIG EYLEKII PLVVKFCNVDDDELREYCIQAFESFVRRCPKEV YPHVSTIINI 300
            ISRQAGHRIG EYLEKII PLVVKFCNVDDDELREYCIQAFESFVRRCPKEV YPHVSTIINI
Sbjct:      241 ISRQAGHRIG EYLEKII PLVVKFCNVDDDELREYCIQAFESFVRRCPKEV YPHVSTIINI 300

Query:      301 CLKYLTYPDN YNYDDEDEDE NAMDADGGDDDD QGSDDDEYS DDDMSWKVR RAAAKCLDAV 360
            CLKYLTYPDN YNYDDEDEDE NAMDADGGDDDD QGSDDDEYS DDDMSWKVR RAAAKCLDAV
Sbjct:      301 CLKYLTYPDN YNYDDEDEDE NAMDADGGDDDD QGSDDDEYS DDDMSWKVR RAAAKCLDAV 360

Query:      361 VSTRHEMLPE FYKTVSPALISRFKEREENV KADV FHAYLS LLKQTRPVQS WLCDDPDAMEQ 420
            VSTRHEMLPE FYKTVSPALISRFKEREENV KADV FHAYLS LLKQTRPVQS WLCDDPDAMEQ
Sbjct:      361 VSTRHEMLPE FYKTVSPALISRFKEREENV KADV FHAYLS LLKQTRPVQS WLCDDPDAMEQ 420
```

Query: 421 GETPLTMLQSQVFNIVKALHKQMKESVKTRQCCFNMLTELNVNLPALTQHIPPVLVPGI 480
 GETPLTMLQSQVFNIVKALHKQMKESVKTRQCCFNMLTELNVNLPALTQHIPPVLVPGI
 Sbjct: 421 GETPLTMLQSQVFNIVKALHKQMKESVKTRQCCFNMLTELNVNLPALTQHIPPVLVPGI 480

Query: 481 IFSLNDKSSSSNLKIDALSCLYVILCNHSPQVFHHPVQALVPPVACVGDFFYKITSEAL 540
 IFSLNDKSSSSNLKIDALSCLYVILCNHSPQVFHHPVQALVPPVACVGDFFYKITSEAL
 Sbjct: 481 IFSLNDKSSSSNLKIDALSCLYVILCNHSPQVFHHPVQALVPPVACVGDFFYKITSEAL 540

Query: 541 LVTQQLVKVIRPLDQPSFDPATPYIKDLFTCTIKRLKAADIDQEVKERAISCMGQIICNL 600
 LVTQQLVKVIRPLDQPSFDPATPYIKDLFTCTIKRLKAADIDQEVKERAISCMGQIICNL
 Sbjct: 541 LVTQQLVKVIRPLDQPSFDPATPYIKDLFTCTIKRLKAADIDQEVKERAISCMGQIICNL 600

Query: 601 GDNLGSDLPNTLQIFLERLKNEITRLTTVKALTLIAGSPKIDLRPVLGEGVPILASFLR 660
 GDNLG DL NTLQIFLERLKNEITRLTTVKALTLIAGSPKIDLRPVLGEGVPILASFLR
 Sbjct: 601 GDNLGSDLPNTLQIFLERLKNEITRLTTVKALTLIAGSPKIDLRPVLGEGVPILASFLR 660

Query: 661 KNQRALKLGTLSALDILIKNYSDSLTAAMIDAVLDELPLPILISESDMHVSQMAISFLTTLA 720
 KNQRALKLGTLSALDILIKNYSDSLTAAMIDAVLDELPLPILISESDMHVSQMAISFLTTLA
 Sbjct: 661 KNQRALKLGTLSALDILIKNYSDSLTAAMIDAVLDELPLPILISESDMHVSQMAISFLTTLA 720

Query: 721 KVPYSSLSKISGSILNELIGLVRSPLLQGGALSAMLDFFQALVVTGTNNLGMDLLRMLT 780
 KVPYSSLSKISGSILNELIGLVRSPLLQGGALSAMLDFFQALVVTGTNNLGMDLLRMLT
 Sbjct: 721 KVPYSSLSKISGSILNELIGLVRSPLLQGGALSAMLDFFQALVVTGTNNLGMDLLRMLT 780

Query: 781 GPVYSQSTALTHKQSYYSIAKCVAAALTRACPKEGPAVVGQFIQDVKNRSRSTDSIRLLALL 840
 GPVYSQSTALTHKQSYYSIAKCVAAALTRACPKEGPAVVGQFIQDVKNRSRSTDSIRLLALL
 Sbjct: 781 GPVYSQSTALTHKQSYYSIAKCVAAALTRACPKEGPAVVGQFIQDVKNRSRSTDSIRLLALL 840

Query: 841 SLGEVGHHDLSGQLELKSIVLEAFSSPSEEVKSAASYALGSISVGNLPEYLPFVLQEIT 900
 SLGEVGHHDLSGQLELKSIVLEAFSSPSEEVKSAASYALGSISVGNLPEYLPFVLQEIT
 Sbjct: 841 SLGEVGHHDLSGQLELKSIVLEAFSSPSEEVKSAASYALGSISVGNLPEYLPFVLQEIT 900

Query: 901 SQPKRQYLLHSLKEIISASVVGKPYVENIALLKHCECAEEGTRNVVAECLGKLT 960
 SQPKRQYLLHSLKEIISASVVGKPYVENIALLKHCECAEEGTRNVVAECLGKLT
 Sbjct: 901 SQPKRQYLLHSLKEIISASVVGKPYVENIALLKHCECAEEGTRNVVAECLGKLT 960

Query: 961 IDPETLLPRLKGYLISGSSYARSSVVTAVKFTISDHPQPIDPLKNCIGDFLKTLED PDL 1020
 IDPETLLPRLKGYLISGSSYARSSVVTAVKFTISDHPQPIDPLKNCIGDFLKTLED PDL
 Sbjct: 961 IDPETLLPRLKGYLISGSSYARSSVVTAVKFTISDHPQPIDPLKNCIGDFLKTLED PDL 1020

Query: 1021 NVRRVALVTFNSSAAHNKPSLIRDLLDVLPHLYNETKVRKELIREVEMGPFKHTVDDGLD 1080
 NVRRVALVTFNSSAAHNKPSLIRDLLD+VLP HLYNETKVRKELIREVEMGPFKHTVDDGLD
 Sbjct: 1021 NVRRVALVTFNSSAAHNKPSLIRDLLDVLPHLYNETKVRKELIREVEMGPFKHTVDDGLD 1080

Query: 1081 IRKAAFECCMYTLLDSCDLRLDIFEFLNHVEDGLKDHYDIKMLTFLMLVRLSTLCPSAVLQ 1140
 IRKAAFECCMYTLLDSCDLRLDIFEFLNHVEDGLKDHYDIKMLTFLMLVRLSTLCPSAVLQ
 Sbjct: 1081 IRKAAFECCMYTLLDSCDLRLDIFEFLNHVEDGLKDHYDIKMLTFLMLVRLSTLCPSAVLQ 1140

Query: 1141 RLDRLVEPLRATCTTKVKANSVKQEFKQDELKRSAMRAVAALLTIPEAEKSPMLSEFQS 1200
 RLDRLVEPLRATCTTKVKANSVKQEFKQDELKRSAMRAVAALLTIPEAEKSPMLSEFQS
 Sbjct: 1141 RLDRLVEPLRATCTTKVKANSVKQEFKQDELKRSAMRAVAALLTIPEAEKSPMLSEFQS 1200

Query: 1201 QISSNPELAAIFESIQRDSSSTNLESMDTS 1230
 QISSNPELAAIFESIQRDSSSTNLESMDTS
 Sbjct: 1201 QISSNPELAAIFESIQRDSSSTNLESMDTS 1230

Pedant information for DKFZphtes3_22g2, frame 2

Report for DKFZphtes3_22g2.2

[LENGTH] 1230
 [MW] 136376.58
 [pI] 5.52
 [HOMOL] TREMBL:RND6711_1 gene: "tip120"; product: "TIP120"; Rattus norvegicus mRNA for
 TIP120, complete cds. 0.0
 [KW] TRANSMEMBRANE 1
 [KW] LOW_COMPLEXITY 5.28 %

SEQ MASASYHISNLEKMTSSDKDFRMTNDLMTELQKDSIKLDDDSERKVVKMILKLEDK
 SEG
 PRD cccccchhhhhhhhhccccceeeehhhhhhhhhccccccccchhhhhhhhhhhhhcc
 MEM

SEQ NGEVQNLAVKCLGPLVSKVKEYQVETIVDTLCTNMLSDKEQLRDISSIGLKTIVIGELPPA
 SEGxxxx
 PRD cccccceeeeeeeehhhhhhhhhhhhhccccchhhhhccccccccchhhhhhhhhhhccccc

MEM
SEQ SSGSALAANVCKKITGRLTSAIAKQEDVSVQLEALDIMADMLSRQGGLLVNFHPSILTCL
SEG xxxxxxxx.....
PRD cccccchhhhhhhccchhhhhhhcccccchhhhhhhhhhhhhhhccceeeccchhhhh
MEM
SEQ LPQLTSPRLAVRKRTIIALGHLVMSGNIVFVDLIEHLLSELSKNDSMSTTRTYIQCIAA
SEG
PRD hccccchhhhhhhhhhhheeeccceehhhhhhhhhhhcccccchhhhhhhhhhh
MEMMMMMMMMMMMMMMM.....
SEQ ISRQAGHRIGEYLEKIIPLVVKFCNVDDDELREYCIQAFESFVRRCPKEVYPHVSTIINI
SEG
PRD hhhhhccccccccchhhhhhhheeeccchhhhhhhhhhhhhhhcccccceccchhhhh
MEM
SEQ CLKYLTYPNPNYDDEDEDENAMDADGGDDDDQGSDDDEYSDDDMSWKVRRAAKCLDAV
SEG xx.....
PRD hhhhhccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhh
MEM
SEQ VSTRHEMLPEFYKTVSPALISRFKEREENVKADVHFAYLSLLKQTRPVQSWLCDPDAMEQ
SEG
PRD hhhccceeeccccc
MEM
SEQ GETPLTMLSQVFNIVKALHKQKESVKTRQCCFNMLTELNVNLPGALTQHIPVLVPGI
SEG
PRD ccccchhhcccccceeeccce
MEM
SEQ IFSLNDKSSSSNLKIDALSCLYVILCNHSPQVFHPPHVALVPPVAVCVGDPFYKITSEAL
SEG xxxxxxxxxxxxxxxx.....
PRD eeecccccchhhhhhhheeecccccceeeccceeecccccchhhhhhh
MEM
SEQ LVTQQLVKVIRPLDQPSFDPATPYIKDLFTCTIKRLKAADIDQEVKERAISCMGQIICNL
SEG
PRD hhhhhhhhhhhccccccccccccchhhhhhhhhhhhhhhccchhhhhhhhhhhheeecc
MEM
SEQ GDNLGSDLPNTLQIFLERLKNEITRLTTVKALTLIAGSPLKIDLRPVLGEGVPILASFLR
SEG
PRD cccccccccchhhhhhhhhhhhhhhhhhhhhhhheeecccccceeehhhhhhhhhh
MEM
SEQ KNQRALKGLTSLDILIKNYSDSLTAAMIDAVLDELPLISESDMHVSQMAISFLTTLA
SEG
PRD hhhhhhhhhhhhhhhhhhhhhcccccchhhhhhhhhhhhhcccccchhhhhhhhhhhhhhh
MEM
SEQ KVYPSSLSKISGISILNELIGLVRSPLLQGGALSAMLDFFQALVVTGTNNLGYMDLRLMT
SEG
PRD cccccceecchhhhhhhhhhhhhcccccchhhhhhhhhhhheeecccccchhhhhhhhh
MEM
SEQ GPVYSQSTALTHKQSYYSIAKCVAAALTRACKEGPAVVGQFIQDVKNRSTDSIRLLALL
SEG
PRD cccccccccchhhhhhhhhhhhhhhhhhhhhcccccchhhhhhhhhhhcccccchhhhhhh
MEM
SEQ SLGEVGHIDLSGQLELKSIVILEAFSSPSEEVKSAASYALGSISVGNLPEYLPFVLQEIT
SEG
PRD hccccccccccccccccceeecccccchhhhhhhhhhhcccccchhhhhhhhh
MEM
SEQ SQPKRQYLLHSLKEI ISSASVVGLKPYVENIWALLLKHCECAEEGTRNVVAECLGKLT
SEG
PRD ccccchhhhhhhhhhhhhcccccchhhhhhhhhhhhhhhcccccceeeccccc
MEM
SEQ IDPETLLPRLKGYLISGSSYARSSVVTAVKFTISDHPQPIDPLKNCIGDFLKTLEDPL
SEG
PRD cccccccccccccccccccccchhhhhhhhhhhcccccchhhhhhhhhhhccccc
MEM
SEQ NVRRVALVTFNSAAHNKPSLIRDLLDTVLPHLYNETKVRKELIREVEMGPFKHTVDDGLD
SEG
PRD cceeeeeeccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccch
MEM

```
SEQ  IRKAAFECCMYTLLDSCLDRLDIFEFLNHVEDGLKDHYDIKMLTFLMLVRLSTLCPSAVLQ
SEG  .....
PRD  hhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  RLDRLVEPLRATCTTKVKANSVKQEFKQDELKRSAMRAVAALLTIPEAEKSPLMSEFQS
SEG  .....
PRD  hhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  QISSNPELAAIFESI QKSSSTNLESMDTS
SEG  .....
PRD  hhhccchhhhhhhhhhhcccccccccccccccccccccccc
MEM  .....
```

(No Prosite data available for DKFZphtes3_22g2.2)

(No Pfam data available for DKFZphtes3_22g2.2)

DKFZphtes3_22n13

group: testes derived

DKFZphtes3_22n13 encodes a novel 677 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

dJ1042K10.3, complete

Sequenced by LMU

Locus: /map="22q13.1-13.2"

Insert length: 3353 bp

Poly A stretch at pos. 3315, polyadenylation signal at pos. 3298

```
1 ATGGAACCACT TATCCCCACT GCCAAGTCCA CCCCCACACT CATTAAAGCAA
51 AGCCAAACCA AGTCTGCCAG TGAGAAGTCA CAGCGCAGCA AGAAGGCCAA
101 GGAGCTGAAG CCAAAGGTGA AGAAGCTCAA GTACCACCAG TACATCCCCC
151 CGGACCAGAA GCAGGACAGG GGGGCACCCC CCATGGACTC ATCCTACGCC
201 AAGATCCTGC AGCAGCAGCA GCTCTTCCTC CAGCTGCAGA TCCTCAACCA
251 GCAGCAGCAG CAGCACCACA ACTACCAGGC CATCCTGCCT GCGCCGCCAA
301 AGTCAGCAGG CGAGGCCCTG GGAAGCAGCG GGACCCCCC AGTACGCAGC
351 CTCTCCACTA CCAATAGCAG CTCCAGCTCG GCGCCCCCTG GGCCCTGTGG
401 GCTGGCAGCT CAGAACAGCA CCTCACTGAC TGGCAAGCCG GGAGCCCTGC
451 CGGCCAACCT GGACGACATG AAGGTGGCAG AGCTGAAGCA GGAGCTGAAG
501 TTGCGATCAC TGCCTGTCTC GGGCACCAAA ACTGAGCTGA TTGAGCGCCT
551 TCGAGCCTAT CAAGACCAAA TCAGCCCTGT GCCAGGAGCC CCCAAGGCCC
601 CTGCGGCCAC CTCTATCCTG CACAAGGCTG GCGAGGTGGT GGTAGCCTTC
651 CCAGCGGCCG GGCTGAGCAC GGGGCCAGCC CTGGTGGCAG CAGGCCCTGGC
701 TCCAGCTGAG GTGGTGGTGG CCACGGTGGC CAGCAGTGGG GTGGTGAAGT
751 TTGGCAGCAC GGGCTCCACG CCCCCCGTGT CTCCACCCC CTGCGAGCGC
801 TCACTGCTAG CACGGGCGGA TGAAAACCTC ACCCCCGGGG ACACCTTTGG
851 TGAGATGGTG ACATCACCTC TGACGCAGCT GACCTGTCAG GCCTCGCCAC
901 TGCAGATCCT CGTGAAGGAG GAGGGCCCCC GGGCCGGGTC CTGTTGCCTG
951 AGCCCTTGGG GCGGGGCGGA GCTAGAGGGG GCGGACAAGG ACCAGATGCT
1001 GCAGGAGAAA GACAAGCAGA TCGAGGCGCT GACGCGCATG CTCCGGCAGA
1051 AGCAGCAGCT GGTGGAGCGG CTCAAGCTGC AGCTGGAGCA GGAGAAGCGA
1101 GCGCCAGCAG CCGCCCCCGC CCGCCGCCCT CTCGGCACCC CCGTGAAGCA
1151 GGAGAACAGC TTCTCCAGCT GCCAGCTGAG CCAGCAGCCC CTGGGCCCCG
1201 CTCACCCATT CAACCCAGC CTGGCGGCCC CAGCCACCAA CCACATAGAC
1251 CCTTGTGCTG TGGCCCCAGG GCCCCCGTCC GTGGTGGTGA AGCAGGAAGC
1301 CTTGACAGCT GAGCCCCAGC CGGTCCCCGC CCCCCAGTTG CTCTGGGGC
1351 CTCAGGGCCC CGGCCTCATC AAGGGGGTTG CACCTCCAC CTCATCACC
1401 GACTCCACAG GGACCCACCT TGTCTCACC GTGACCAATA AGAATGCAGA
1451 CAGCCCTGGC CTGTCCAGTG GGAGCCCCCA GCAGCCCTCG TCCCAGCCTG
1501 GCTCTCCAGC GCCTGCCCCC TCTGCCAGA TGGACCTGGA CAGCCACTG
1551 CAGCCCTCTT TTGGGACCCC CACTTCTCTG CTGAAGAAGG AACCACCTGG
1601 CTATGAGGAA GCCATGAGCC AGCAGCCCCA ACAGCAGGAA AATGGTTCTT
1651 CAAGCCAGCA GATGGACGAC CTGTTTGACA TTCTCATTCA GAGCGGAGAA
1701 ATTTCAAGCA ATTTCAAGGA GCCGCCATCC CTGCCAGGGA AGGAGAAGCC
1751 ATCCCGAAG ACAGTCTGTG GGTCCCCCCT GGCAGCACAG CCATCACCTT
1801 CTGCTGAGCT CCCCCAGGCT GCCCCACCTC CTCCAGGCTC ACCCTCCCTC
1851 CCTGGACGCC TGGAGGACTT CCTGGAGAGC AGCACGGGGC TGCCCTTGCT
1901 GACAGTGGG CATGACGGGC CAGAGCCCCT TTCCCTCATT GACGACCTCC
1951 ATAGCCAGAT GCTGACGAGC ACTGCCATCC TGGACCACCC CCGTCAACCC
2001 ATGGACACCT CGGAATTGCA CTTTGTTCCT GAGCCAGCA GCACCATGGG
2051 CCTGGACCTG GCTGATGGCC ACCTGGACAG CATGGAAGTG CTGGAGCTGT
2101 CGTCAGGTGG TCCCGTGCTG AGCCTAGCCC CCCTCAGCAC CACAGCCCCC
2151 AGCCTCTTCT CCACAGACTT CCTCGATGGC CATGATTGCG AGCTGCAGTG
2201 GGATTCTCTG TTGTAGCTCT CTGGCTCAAG ACGGGGTGGG GAAGGGGCTG
2251 GGAGCCAGGG TACTCCAATG CGTGGCTCTC CTGCGTGATT CGGCCTCTCC
2301 ACATGGTTGT GAGTCTTGAC AATCAGAGCC CTGTCTTTT CCCTTCCCTG
2351 GGAGGCTAGA ACAGAGAAGC CTTACTCTCT GGTTCAGTGC CAGCAGGGC
2401 AGAGGAGAGC AGCTGTCAAG AAGCAGCCCT GGCTCTCAGC CTGGGGTTTT
2451 GGACACACGG TCAGGGTCAG GGCCATTTC AATTGACCTC CTTTTTTGAG
2501 GTCAGGGGGC ACTGTCTGTC TGGCTACAAT TTGGCTAAGG TAGGTGAAGC
2551 CTGGCCAGGC GGGAGGCTTC TCTCTGACC CAGGGCTGAG ACAGGTTAAG
2601 GGGTGAATCT CCTTCTTTC TCTCCCTGCT TTGCTGTGAA GGGAGAAATT
2651 AGCCTGGGCC TCTACCCCTT ATTCCTGTG TCTGCCAACC CCAGGATCCC
2701 AGGGTCCCT GCCATTTTAG TGTCTTGGTG TAGTGTAAAC ATTAGTGGT
2751 TGGTGGCAAC AATTTTATGT ACAGGTGTAT ATACCTCTAT ATTATATATC
2801 GACATACATA TATATTTTGG GGGGGGGGCG GACAGGAGAT GGGTGCAACT
```

```

2851 CCCTCCCATC CTACTCTCAC AGAAGGGCCT GGATGCAAGG TTACCCTTGA
2901 GCTGTGTGCC ACAGTCTGGT GCCCAGTCTG GCATGCAGCT ACCCAGGCCC
2951 ACCCATCAGC TGTGATTGAC ATGTAGGTAC CCTGCCACGG CCTATGCCCC
3001 ACCTGCCCTG CTTCCTGGCT CCTATCAGT GCCATGAGGG CAGAGGTGCT
3051 ACCTGGCCTT CCTGCCAGGA GCTCTCCACC CACTCACATT CCGTCCCCGC
3101 CGCCTCACAT CAGCCAGCGT GGCCCTAGGA CAGGAGGAGC TTCGGGCCCCA
3151 GCTTCACCTT GCGGTGGGGC TGAGGGGTGG CCATCTCCTG CCCTGGGGCC
3201 ACTGGCTTCA CATTCTGGGC TGAATCATAG GGGAGTAGGG GTGGAGTCAC
3251 CAAAACACAGT GCTGGGACAA AGATGGGGAA GGTGTGTGAA CTTTTTAAAA
3301 TAAACACAAA AACACAGGAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3351 AAG

```

BLAST Results

Entry HS1042K10 from database EMBL:
 Human DNA sequence from clone 1042K10 on chromosome 22q13.1-13.2.
 Contains the ADSL gene for Adenylosuccinate lyase (EC 4.3.2.2, Adenylosuccinase, ASL) and 4 novel genes (one with probable rabGAP domains and Src homology domain 3). Contains ESTs, STSs, GSSs and a putative CpG island.
 Score = 7997, P = 0.0e+00, identities = 1617/1645
 7 exons

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 183 bp to 2213 bp: peptide length: 677
 Category: similarity to unknown protein
 Classification: unclassified

```

1 MDSSYAKILQ QQQLFLQLQI LNQQQQQHNN YQAILPAPPK SAGEALGSSG
51 TPPVRSLSST NSSSSSGAPG PCGLARQNST SLTGKPGALP ANLDDMKVAE
101 LKQELKLRSL PVSGETKTELI ERLRAYQDQI SPVPGAPKAP AATSILHKAG
151 EVVVAEPFAAR LSTGPALVAA GLAPAEVVVA TVASSGVVKE GSTGSTPPVS
201 PTPSERSLLS TGDENSTPGD TFGEMVTSPL TQLTLQASPL QILVKEEGPR
251 AGSCCLSPGG RAELEGRDKD QMLQEKDKQI EALTRMLRQK QQLVERLKLQ
301 LEQEKRAQQP APAPAPLGTP VKQENSFSSC QLSQQPLGPA HPFNPSLAAP
351 ATNHIDPCAV APGPPSVVVK QEALQPEPEP VPAPQLLLGP QGPGLIKGVA
401 PPTLITDSTG THLVLTVTNK NADSPGLSSG SPQQPSSQPG SPAPAPSAQM
451 DLEHPLQPLF GTPTSLLKKE PPGYEEAMSQ QPKQQENGSS SQQMDDLFDI
501 LIQSGEISAD FKEPPSLPGK EKPSPKTVCG SPLAAQSPSP AELPQAAPPP
551 PGSPSLPGRL EDFLESSTGL PLLTSGHDGP EPLSLIDDLH SQMLSSTAIL
601 DHPPSPMDTS ELHFVPEPSS TMGLDLADGH LDSMDWLELS SGGPVLSLAP
651 LSTTAPSLFS TDFLDGHDLO LHWDSCL

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_22n13, frame 3

TREMBL:HS1042K10_6 gene: "dJ1042K10.3"; product: "dJ1042K10.3 (novel protein)"; Human DNA sequence from clone 1042K10 on chromosome 22q13.1-13.2. Contains the ADSL gene for Adenylosuccinate lyase (EC 4.3.2.2, Adenylosuccinase, ASL) and 4 novel genes (one with probable rabGAP domains and Src homology domain 3). Contains ESTs, STSs, GSSs and a putative CpG island., N = 1, Score = 1285, P = 4.9e-131

TREMBL:CEUK06A9_3 gene: "K06A9.1a"; Caenorhabditis elegans cosmid K06A9., N = 2, Score = 149, P = 1.3e-09

TREMBLNEW:SSI132828_1 product: "p210 protein"; Spermatozoopsis similis mRNA for p210 protein, partial, N = 1, Score = 171, P = 2.8e-09

>TREMBL:HS1042K10_6 gene: "dJ1042K10.3"; product: "dJ1042K10.3 (novel protein)"; Human DNA sequence from clone 1042K10 on chromosome 22q13.1-13.2. Contains the ADSL gene for Adenylosuccinate lyase (EC

HSPs :

Query:	435	PSSQPGSPAPAPSAQMDEHPLQLFLGFTPTSLLKKEPPGYEEAMSSQQPKQEQNGSSSQQM	494
		PSSQPGSPAPAPSAQMDEHPLQLFLGFTPTSLLKKEPPGYEEAMSSQQPKQEQNGSSSQQM	
Sbjct:	1	PSSQPGSPAPAPSAQMDEHPLQLFLGFTPTSLLKKEPPGYEEAMSSQQPKQEQNGSSSQQM	60
Query:	495	DDLFDLILIQSGEISADFKEPPSLPGKEKPSPKTVCGSPLAAQPSAELPQAAPPPPGSP	554
		DDLFDLILIQSGEISADFKEPPSLPGKEKPSPKTVCGSPLAAQPSAELPQAAPPPPGSP	
Sbjct:	61	DDLFDLILIQSGEISADFKEPPSLPGKEKPSPKTVCGSPLAAQPSAELPQAAPPPPGSP	120
Query:	555	SLPGRLEDLFLESSTGLPLLTSGHDGPEPLSLIDDLHSQMLSSTAILDHPSPMDTSELHF	614
		SLPGRLEDLFLESSTGLPLLTSGHDGPEPLSLIDDLHSQMLSSTAILDHPSPMDTSELHF	
Sbjct:	121	SLPGRLEDLFLESSTGLPLLTSGHDGPEPLSLIDDLHSQMLSSTAILDHPSPMDTSELHF	180
Query:	615	VPEPSSTMGLDLADGHLDSMDWLELSSGGPVLSLAPLSTTAPSLFSTDFLDGHDLQLHWD	674
		VPEPSSTMGLDLADGHLDSMDWLELSSGGPVLSLAPLSTTAPSLFSTDFLDGHDLQLHWD	
Sbjct:	181	VPEPSSTMGLDLADGHLDSMDWLELSSGGPVLSLAPLSTTAPSLFSTDFLDGHDLQLHWD	240
Query:	675	SCL 677	
		SCL	
Sbjct:	241	SCL 243	

Pedant information for DKFZphtes3 22n13, frame 3

Report for DKFZphtes3 22n13.3

```
[LENGTH]          677
[MW]               70743.01
[pI]              4.93
[HOMOL]           TREMBL:HS1042K10.6 gene: "dJ1042K10.3"; product: "dJ1042K10.3 (novel protein)";
Human DNA sequence from clone 1042K10 on chromosome 22q13.1-13.2. Contains the ADSL gene for
Adenylosuccinate lyase (EC 4.3.2.2, Adenylosuccinase, ASL) and 4 novel genes (one with
probable rabGAP domains and Src homology domain 3). Contains ESTs, STSS, GSSs and a putative
CpG island. 1e-111
[KW]              TRANSMEMBRANE 1
[KW]              LOW COMPLEXITY 21.57 %
[KW]              COILED COIL 4.58 %
```

```

SEQ      MDSSYAKILQOOQLFLQLQILNQOOQHNNYQAILPAPPKSAGEALGSSGTPPVRSLSSTT
SEG      . . . . . xxxxxxxxxxxxxxxxxxxxxxx . . . . . xxxxxx
PRD      ccchhhhhhhhhhhhhhhhhhhhhhhhhhhccceeeeeecccccccccccccccccccccc
COILS    . . . . .
MEM      . . . . .

SEQ      NSSSSSGAPGPCGLARQNSTSLTGKPGALPANLDDMKVAELKQELKRLSLPVSGTKTELI
SEG      xxxxxx.
PRD      cccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhcccccchhhh
COILS    . . . . .
MEM      . . . . .

SEQ      ERLRAYQDQISPPVGAPKAPAATSIILHKAGEVVVAFPAARLSTGPALVAAGLAPAEVVVA
SEG      . . . . . xxxxxxxxxxxxxxxxxxxxxx
PRD      hhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS    . . . . .
MEM      . . . . . MMMMMMMMMMMMMMMMMMMMMMMMMM

SEQ      TVASSGVVFGSTGSTPPVSTPSPERSLLSTGDENSTPGDTFGEMVTSPLTQLTLQASPL
SEG      xxxxxxxxx . xxxxxxxxxxxxxxxxxxxxx
PRD      eeeeeccccccccccccccccccccccccceeecccccccccccccccccccccccccccccc
COILS    . . . . .
MEM      M . . . . .

SEQ      QILVKEGPRAGSCCLSPGGRAELEGRDKDQMLQEKDKQIEALTRMLRQKQQLVERLKLQ
SEG      . . . . .
PRD      eeeeeccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    . . . . . CCCCCCCCCCCCCCCCCCCCCCCC
MEM      . . . . .

SEQ      LEQEKRAOQAPAPAPLPTGPKVQENSFSSCOLSOOPLGPAHPFNPSLAAPATNHIDPCA

```

```
SEG .....XXXXXXXXXX.....
PRD hhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS CCCCCC.....
MEM .....

SEG APGPSPVVVKQEQALQPEPEPVPAQQLLGPQGPGLIKGVAPPTLITDSTGTHLVLTVTNK
SEG .....XXXXXXXXXX.....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....
MEM .....

SEG NADSPGLSSGSPQQPSSQPGSPAPAPSAQMDLEHPLQPLFGTPTSLKKEPPGYEEAMSQ
SEG .....XXXXXXXXXX.....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....
MEM .....

SEG QPKQENGSSSQMDDLFDILIQSGEISADFKPEPSPGKEKPSPKTVCGSPLAAQPSPS
SEG .....XXXXXXXXXX.....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....
MEM .....

SEG AELPQAAPPPPGSPSLPGRLEDFLESSTGLPLLTSGHDGPEPLSLIDDLHSQMLSSTAIL
SEG .....XXXXXXXXXX.....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....
MEM .....

SEG DHPPSPMDTSELHFVPEPSSTMGLDLADGHLDSMDWLELSSGGPVLSLAPLSTTAPSLFS
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....
MEM .....

SEG TDFLDGHDLQLHWDSC
SEG .....
PRD ccccccccccccccccccc
COILS .....
MEM .....
```

(No Prosite data available for DKFZphtes3_22n13.3)

(No Pfam data available for DKFZphtes3_22n13.3)

DKFZphtes3_23111

group: intracellular transport and trafficking

DKFZphtes3_23111 encodes a novel 186 amino acid protein nearly identical to mouse ADP-ribosylation-like factor homolog 6 (Arl6).

Protein secretion through the endoplasmic reticulum and the Golgi vesicular trafficking system is initiated by the binding of ADP-ribosylation factors (ARFs) to donor membranes, leading to recruitment of cocatomer, bud formation, and eventual vesicle release. ARFs are approximately 20-kDa GTPases that are active with bound GTP and inactive with GDP bound. The novel protein contains an ATP/GTP-binding site motif A (P-loop) and seems to be a novel ARF. It seems to have an important role in vesicular transport and vesicular trafficking.

The new protein can find application in modulating vesicle transport and trafficking in cells.

nearly identical to mouse Arl6, ADP-ribosylation-like factor homolog

start at Bp 15 matches kozak consensus ANNatgG

Sequenced by LMU

Locus: unknown

Insert length: 717 bp

Poly A stretch at pos. 689, no polyadenylation signal found

```

1  ATTTGAATCA CATTATGGGA TTGCTAGACA GACTTTCAGT CTTGCTTGGC
51 CTGAAGAAGA AGGAGGTTC A TGTTTTGTGC CTTGGGCTAG ATAATAGTGG
101 CAAAACGACG ATCATTAAAC AACTTAAACC TTCAAATGCT CAATCTCAAA
151 ATATCCTTCC AACAAATAGGA TTCAGCATAG AGAAATTCAA ATCATCCAGT
201 TTGTCATTTA CAGTGTTTGA CATGTCAGGT CAAGGAAGAT ACAGAAATCT
251 CTGGGAACAC TATTATAAAG AAGGCCAAGC TATTATTTT GTCATTTGATA
301 GTAGTGATAG ATTAAGAATG GTTGTGGCCA AAGAAGAAGT CGATACTCTT
351 CTGAATCATC CAGATATTAA ACACCGTCGA ATTCCAATCT TATTCCTTGC
401 AAATAAAATG GATCTTAGAG ATGCAGTGAC ATCTGTAAAA GTGTCTCAGT
451 TGCTGTGTTT AGAGAACATC AAAGATAAAC CCTGGCATAT TTGTGCTAGT
501 GATGCCATAA AAGGAGAAGG CTTGCAAGAA GGTGTAGACT GGCTTCAAGA
551 TCAGATCCAG ACTGTGAAGA CATGAAAAGA TAATAGTTGG AAACCTCAGC
601 AATTTTCAAT TCAAGGAATC TATCTAAGAC AAATAGAATA CATTTTGTAA
651 AAGATGTTTA TGCATCAAAA AATATAATTT TCTGCTTGCA AAAAAAAAAA
701 AAAAAAAAAA AAAAAAG

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 15 bp to 572 bp; peptide length: 186
 Category: strong similarity to known protein
 Classification: Intracellular transport and traffic
 Prosite motifs: ATP_GTP_A (24-32)

```

1  MGLLDRLSVL LGLKKKEVHV LCLGLDNSGK TTIINKLKPS NAQSQNILPT
51 IGFSIEKFKS SLSFTVFDM SGQGRYRN LW EHYKQEI IFVIDSSDRL
101 RMVVAKEELD TLLNHPDIKH RRIPIFFAN KMDLRDAVTS VKVSQLLCLE
151 NIKDKPWHIC ASDAIKGEGL QEGVDWLQDQ IQTVKT

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_23111, frame 3

TREMBL:AF031903_1 gene: "Arl6"; product: "ADP-ribosylation-like factor homolog ARL6"; Mus musculus ADP-ribosylation-like factor homolog ARL6 (Arl6) mRNA, complete cds., N = 1, Score = 923, P = 1.1e-92

TREMBL:CEC38D4_5 gene: "C38D4.8"; Caenorhabditis elegans cosmid C38D4, N = 1, Score = 418, P = 3.6e-39

PIR:S66337 ADP-ribosylation factor 1 - Chlamydomonas reinhardtii, N = 1, Score = 373, P = 2.1e-34

SWISSPROT:ARF1_CHLRE ADP-RIBOSYLATION FACTOR 1., N = 1, Score = 372, P = 2.7e-34

>TREMBL:AF031903_1 gene: "Arl6"; product: "ADP-ribosylation-like factor homolog ARL6"; Mus musculus ADP-ribosylation-like factor homolog ARL6 (Arl6) mRNA, complete cds.
Length = 186

HSPs:

Score = 923 (138.5 bits), Expect = 1.1e-92, P = 1.1e-92
Identities = 178/186 (95%), Positives = 184/186 (98%)

```
Query:      1 MGLDLRLSVLLGLKKKEVHVLCGLDNSGKTTIINKLKPSNAQSQNILPTIGFSIEKFKS 60
             MGLDLRLS LLGLKKKEVHVLCGLDNSGKTTIINKLKPSNAQSQ+I+PTIGFSIEKFKS
Sbjct:      1 MGLDLRLSGLLGLKKKEVHVLCGLDNSGKTTIINKLKPSNAQSQDIVPTIGFSIEKFKS 60

Query:      61 SSLSFTVFDMMSGQGRYRNLWEHYKQGAIIFVIDSSDRLRMVVAKEELDTLLNHPDIKH 120
             SSLSFTVFDMMSGQGRYRNLWEHYK+GQAIIFVIDSSD+LRMVVAKEELDTLLNHPDIKH
Sbjct:      61 SSLSFTVFDMMSGQGRYRNLWEHYKDGQAIIFVIDSSDKLRMVVAKEELDTLLNHPDIKH 120

Query:      121 RRIPILFFANKMDLRDAVTSVKVSQLLCLENIKDKPWHICASDAIKGEGVQGVVDWLQDQ 180
             RRIPILFFANKMDLRD+VTSVKVSQLLCLE+IKDKPWHICASDAIKGEGVQGVVDWLQDQ
Sbjct:      121 RRIPILFFANKMDLRDSVTSVKVSQLLCLESIKDKPWHICASDAIKGEGVQGVVDWLQDQ 180

Query:      181 IQTVKT 186
             IQ VKT
Sbjct:      181 IQAVKT 186
```

Pedant information for DKFZphtes3_23111, frame 3

Report for DKFZphtes3_23111.3

```
[LENGTH]      186
[MW]           21097.69
[pI]           8.72
[HOMOL]        TREMBL:AF031903_1 gene: "Arl6"; product: "ADP-ribosylation-like factor homolog
ARL6"; Mus musculus ADP-ribosylation-like factor homolog ARL6 (Arl6) mRNA, complete cds. 4e-94

[FUNCAT]       30.08 organization of golgi [S. cerevisiae, YDL192w] 1e-36
[FUNCAT]       06.10 assembly of protein complexes [S. cerevisiae, YDL192w] 1e-36
[FUNCAT]       08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL192w]
1e-36
[FUNCAT]       30.09 organization of intracellular transport vesicles [S. cerevisiae,
YDL137w] 2e-36
[FUNCAT]       06.07 protein modification (glycosylation, acylation, myristylation,
palmitoylation, farnesylation and processing) [S. cerevisiae, YBR164c] 2e-32
[FUNCAT]       30.03 organization of cytoplasm [S. cerevisiae, YBR164c] 2e-32
[FUNCAT]       03.22 cell cycle control and mitosis [S. cerevisiae, YMR138w] 4e-19
[FUNCAT]       30.04 organization of cytoskeleton [S. cerevisiae, YMR138w] 4e-19
[FUNCAT]       r general function prediction [M. jannaschii, MJ1339] 2e-05
[FUNCAT]       30.02 organization of plasma membrane [S. cerevisiae, YHR005c] 4e-05
[FUNCAT]       03.07 pheromone response, mating-type determination, sex-specific proteins
[S. cerevisiae, YHR005c] 4e-05
[FUNCAT]       10.05.07 g-proteins [S. cerevisiae, YHR005c] 4e-05
[FUNCAT]       08.13 vacuolar transport [S. cerevisiae, YKR014c] 2e-04
[FUNCAT]       08.19 cellular import [S. cerevisiae, YKR014c] 2e-04
[FUNCAT]       06.04 protein targeting, sorting and translocation [S. cerevisiae, YKR014c]
2e-04
[FUNCAT]       03.04 budding, cell polarity and filament formation [S. cerevisiae, YFL005w]
4e-04
[BLOCKS]      BL01288C
[BLOCKS]      BL01020C SAR1 family proteins
[BLOCKS]      BL01019C ADP-ribosylation factors family proteins
```

```

[BLOCKS]      BL01019B ADP-ribosylation factors family proteins
[BLOCKS]      BL01019A ADP-ribosylation factors family proteins
[SCOP]        dlas3_2 3.29.1.4.12 Transducin (alpha subunit), insertion domai 2e-45
[SCOP]        dlmh1_ 3.29.1.4.2 Rac1 (Human (Homo sapiens) 2e-46
[SCOP]        d5p21_ 3.29.1.4.1 cH-p21 Ras protein (human (Homo sapiens) 5e-37
[SCOP]        dlhura_ 3.29.1.4.8 ADP-ribosylation factor 1 (ARF1) [human (Hom 4e-61
[SCOP]        dla2kc_ 3.29.1.4.5 Ran Nuclear transport factor-2 (NTF2) [Do 4e-33
[PIRKW]       glycoprotein 2e-33
[PIRKW]       monomer 3e-31
[PIRKW]       P-loop 2e-35
[PIRKW]       lipoprotein 2e-33
[PIRKW]       GTP binding 2e-35
[SUPFAM]      ADP-ribosylation factor 2e-35
[PROSITE]     ATP_GTP_A 1
[PFAM]        ADP-ribosylation factors (Arf family) (contains ATP/GTP binding P-loop)
[KW]          Alpha_Beta
[KW]          3D
[KW]          LOW_COMPLEXITY 5.91 %

```

```

SEQ  MGLLDRLSVLLGLKKKEVHVLCGLDMSGKTTIINKLKPSNAQSQNILPTIGFSIEKFKS
SEG  ..xxxxxxxxxxxxx.....
lhurA .....CCCCEEEEETTTTCHHHHHHHHCCCCEEEE--EETTTEEEEEEEEE

SEQ  SLSFTVFDMMSGQGRYRNLWEHYHKEGQAIIFVIDSSDRLRMVVAKEELDTLLNHPDIKH
SEG  .....
lhurA TTTEEEEEETTTTTCCHHHHHHHHCEEEEEETTTTTHHHHHHHHHHHHHHTTTT--

SEQ  RRIPILFFANKMDLRDAVTSVKVSQLLCLENIKDKPWHICASDAIKGEGLQEGVDWLQDQ
SEG  .....
lhurA TTTEEEEEETTTTTCCHHHHHHHHCGGGTTTCEEEECBTTTBTTHHHHHHHHHHH

SEQ  IQTVKT
SEG  .....
lhurA HHHHC.

```

Prosite for DKFZphtes3_23111.3

PS00017 24->32 ATP_GTP_A PDOC00017

Pfam for DKFZphtes3_23111.3

```

HMM_NAME      ADP-ribosylation factors (Arf family) (contains ATP/GTP binding P-loop)
HMM            *GMgWfsIFrkMWGLWNKEMRILMLGLDNAGKTTILYMLKlgE..IVTTI
               MG++ ++ ++GL +KE+++L LGLDN+GKTTI+++LK+ ++
Query          1 -MGLLDRLSVLLGLKKKEVHVLCGLDMSGKTTIINKLKPSNAQSQNIL 48
HMM            PTIGFNVETVeYKNIKFNVWDVGGQdsIRPYWRHYYPNTDGIWVVDsaD
               PTIGF +E+ + ++F+V+D GQ + R +W HYY + ++II+V+DS+D
Query          49 PTIGFSIEKFKSSLSFTVFDMMSGQGRYRNLWEHYHKEGQAIIFVIDSSD 98
HMM            RDRMeEaKqELHaMLNEEEL..rDAPILIFANKQDLpGAMSesEIREaLG
               R RM AK+EL+ +LN+ ++ R+ P+L FANK DL++A+++ ++L
Query          99 RLRMVVAKEELDTLLNHPDIKHRRIPILFFANKMDLRDAVTSVKVSQLLC 148
HMM            LHeIRCnRPWYIQMCCAVtGEGLYEGMDWLSNYInkRkK*
               L++I+ + PW+I +++A++GEGLE+G DWL ++I+ K
Query          149 LENIK-DKPWHICASDAIKGEGLQEGVDWLQDQIQTVKT 186

```

DKFZphtes3_23n19

group: testes derived

DKFZphtes3_23n19 encodes a novel 387 amino acid protein with similarity to rat protein kinase C-interacting RBCC protein 1.

The novel protein contains not the RING-B box-coiled coil (RBCC) motif of RBCC protein 1, and thus is not a member of this subgroup of RING finger proteins.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to rat protein kinase C-interacting RBCC protein 1

start at Bp 209 matches kozak consensus PyNNatgG
similarity to of C-terminal part to N-terminus of RBCK1

Sequenced by LMU

Locus: unknown

Insert length: 1579 bp
Poly A stretch at pos. 1535, polyadenylation signal at pos. 1515

```
1 CGGAGACCCT CGGGCCGTGT CCATTGTGG GCAAAGCCAG CGGGGCAGGC
51 TTGGCCAGAG TGCACCACTC GGCGCCGTCC CAGGCCCGAC GCTCTGGGCG
101 CGCCCCGAAC CCCAGGTTTC CGGCCCGTGT TTCCGACCGG CGGAGGGGGG
151 TCAGCGGCCG GATCCACAGG AAGCGCGCTC GGAGGGGTGG GACCCGGCCG
201 GACCCGAGAT GGCGCCGCCA GCGGGCGGGG CGGCGGCGGC GGCTCGGAC
251 TTGGGGCTCG CGCAGTGCT CTGGGCTGTG CACGCCGCGG TGAGGCCGCT
301 GGGCGCGCGG CCAGACGCGG AGGCACAGCT GCGGAGGCTG CAGCTGAGCG
351 CGGACCCCTG GAGGCCTGGG CGCTTCCGGC TGGAGCTGCT GGGCGCGGGA
401 CCTGGGGCGG TTAATTGGA GTGGCCCTG GAGTCAGTTT CCTACACCAT
451 CCGAGGCCCC ACCCAGCACG AGCTACAGCC TCCACCAGGA GGGCCTGGAA
501 CCCTCAGCCT GCACTTCCTC AACCCCTCAGG AAGCTCAGCG GTGGGCAGTC
551 CTAGTCCGAG GTGCCACCGT GGAAGGACAG AATGGCAGCA AGAGCAACTC
601 ACCACCAGCC TTGGGCCCAG AAGCATGCCC TGTCTCCCTG CCCAGTCCCC
651 CGGAAGCCTC CACACTCAAG GGCCCTCCAC CTGAGGCAGA TCTTCTTAGG
701 AGCCCTGGAA ACTTGACGGA GAGAGAAGAG CTGGCAGGGA GCCTGGCCCG
751 GGCTATTGCA GGTGGAGACG AGAAGGGGGC AGCCCAAGTG GCAGCCGTCC
801 TGGCCAGCA TCGTGTGGCC CTGAGTGTTT AGCTTCAGGA GGCCTGCTTC
851 CCACCTGGCC CCATCAGGCT GCAGGTGACA CTGAAGACG CTGCCTCTGC
901 CGCATCCGCC GCGTCCTCTG CACACGTTGC CCTGCAGGTC CACCCCCACT
951 GCACTGTTGC AGCTCTCCAG GAGCAGGTGT TCTCAGAGCT CGGTTTCCCG
1001 CCAGCCGTGC AACGCTGGGT CATCGGACGG TGCCCTGTGTG TGCCCTGAGCG
1051 CAGCCTTGCC TCTTACGGGG TTCGGCAGGA TGGGGACCCG GCTTTCTCTT
1101 ACTTGCTGTG AGCTCCTCGA GAAGCCCCAG CCACAGGACC TAGCCCTCAG
1151 CACCCCCAGA AGATGGACGG GGAACCTGGA CGCTTGTTTC CCCCATCATT
1201 GGGGCTACCC CCAGGCCCCC AGCCAGCTGC CTCCAGCCTG CCCAGTCCAC
1251 TCCAGCCCAG CTGGTCTGT CCTTCTGCA CCTTCATCAA TGCCCCAGAC
1301 CGCCCTGGGT GTGAGATGTG TAGCACCCAG AGGCCCTGCA CTGGGGACCC
1351 CTTTGTCTGA GCTTCCACCT AGCAGCCACC AGAGGTTACA AGGGGAGAGT
1401 GGCCCTTCCC TCACAAGTCC GACATCTCCA GGCCCCACT GAACTCGGGG
1451 GACCTCTACT GACTGCTTGC TGGGACAGTC ACCAGGGTTG GGGGGAAGGG
1501 CCACAAAATG AAACCATTA AGACCCTTAA GAGCCAAAAA AAAAAAAAAA
1551 AAAAAAAAAA AAAAAAAAAA AAAAAAAG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 209 bp to 1369 bp; peptide length: 387


```

1  MAPPAGGAAA  AASDLGSAV  LLAVHAAVRP  LGAGPDAAEQ  LRRQLSADP
51  ERPGFRLEL  LGAGPGAVN  EWPLESVSY  IRGPTQHELE  PPPGSGPTLS
101  LHFLPQRE  RAVLVRGAT  VEGQNGSKN  SPPLAGPEAC  PVSLSPPPEA
151  STLKGPPPE  DLPRSPGNT  EREELAGSA  RAIAGGDGK  AAQVAAVLQ
201  HRVALSVQLE  EACFPNGPI  LQVTLDAAS  AASAASHAV  LQVHPHCTV
251  AALQEQVFE  LGFPPAVQR  VIGRCLVPE  RSLASYGVR  QDHPAFLYL
301  SAPREAPAT  PSPQHPQKM  GELGRLFP  LGLPPGPQA  ASSLPSLPQ
351  SWSPCSTPI  NAPDRPGCE  CSTRPCTWD  PLAAAST

```

No BLASTP hits available

```
SEQ      MAPPAGGAAAAASDLGSAAVLLAVHAAVRPLGAGPDAAEQLRRLQLSADPERPGRFRLEL
SEG      .XXXXXXXXXXXXXXXXXXXXXXXXXXXXX.....
PRD      cccccchhhhhhhhhhhhhhhhhhhhhcccccchhhhhhhhhhhcccccceeee
```

738

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 209 bp to 1369 bp; peptide length: 387
 Category: similarity to known protein
 Classification: Cell signaling/communication

```

1 MAPPAGGAAA AASDLGSAAV LLAVHAAVRP LGAGPDAAEQ LRLQLSADP
51 ERPGFRLEL LGAGPGAVNL EWPLESVSYT IRGPTQHELQ PPPGGPGTSL
101 LHFLNPQEAQ RWAVLVRGAT VEGQNGSKSN SPPALGPEAC PVSLPSPPEA
151 STLKGPPPEA DLPRSPGNLT EREELAGSLA RAIAGGDEKG AAQVAVLAQ
201 HRVALSVQLQ EACFPPIR LQVTLEDAAS AASAASSAHV ALQVHPHCTV
251 AALQEQVFSE LGFPPAVQRW VIGRCLCVPE RSLASYGVRQ DGDPAFLYLL
301 SAPREAPATG PSPQHPQKMD GELGRLEFPPS LGLPPGPQPA ASSLPSPQLP
351 SWSCEPCTFI NAPDRPGCEM CSTQRPCTWD FLAAAST

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_23n19, frame 2

PIR:JC5983 protein kinase C-interacting RBCK protein 1 - rat, N = 1,
 Score = 353, P = 2.8e-32

TREMBL:AB011369_1 product: "RBCK2"; Rattus norvegicus mRNA for RBCK2,
 complete cds., N = 1, Score = 353, P = 2.8e-32

TREMBL:U67322_1 gene: "XAP4"; product: "HBV associated factor"; Human
 HBV associated factor (XAP4) mRNA, complete cds., N = 1, Score = 286, P
 = 8.5e-25

TREMBLNEW:AF124663_1 product: "UbcM4 interacting protein 28"; Mus
 musculus UbcM4 interacting protein 28 mRNA, complete cds., N = 1, Score
 = 367, P = 9.3e-34

>TREMBLNEW:AF124663_1 product: "UbcM4 interacting protein 28"; Mus musculus
 UbcM4 interacting protein 28 mRNA, complete cds.
 Length = 498

HSPs:

Score = 367 (55.1 bits), Expect = 9.3e-34, P = 9.3e-34
 Identities = 95/212 (44%), Positives = 129/212 (60%)

```

Query: 175 LAGSLARAIAGGDEKGAQVAVLAQHRLVALSVQLQEACFPPIRLQVTLEDAASAASA 234
      +A SLARA+AGGDE+ A + A LA+ RV L VQ++ P IRL V++EDA
Sbjct: 1 MALSLARAVAGGDEQAQAIKYATWLAEQVRVLRVQVKPEVSPTQDIRLCVSVEDAYM---- 56

Query: 235 ASSAHVALQVHPHCTVAALQEQVFSELGFPPAVQRWVIGRCLCVPERSLASYGVRQDGD 294
      + + L V P TVA+L++ VE + GFPP++Q+WV+G+ L + +L S+G+R++GD
Sbjct: 57 -HTVTIWLTVRPDMTVASLKDVMFLDYGFPPSLQQWVVGQRLARDQETLHSHGIRNRNGDG 115

Query: 295 AFLYLLSAPREAPATGFPSPQHPQK-----MDGELG--RLFPPSLG-LPPG-PQPAASSLP 345
      A+LYLLSA T +PQ Q+ M +LG L S G L P P+P + P
Sbjct: 116 AYL YLLSARN----TSLNPQELQRQRLRMLEDLGFKDLTLQSRGPLEPVLPKPRTNQEP 171

Query: 346 -----SPLQP--SWSCEPCTFINAPDRPGCEMCSTQRPCTW 379
      +P P W CP CTFIN P RPGCEMC RP T+
Sbjct: 172 GQPDAAPEPSPVGVQCGCTFINKPTRPGCEMCRRARPETY 212

```

Pedant information for DKFZphtes3_23n19, frame 2

```
[LENGTH]      387
[MW]           39949.29
[pI]           5.53
[HOMOL]        TREMBLNEW:AF124663_1 product: "UbcM4 interacting protein 28"; Mus musculus
UbcM4 interacting protein 28 mRNA, complete cds. 1e-22
[BLOCKS]       BL00578B
[KW]           Alpha_Beta
[KW]           LOW COMPLEXITY      17.57 %
```

```
SEQ      MAPPAGGAAAAASDLGSAAVLLAVHA AVRPLGAGPDAEAQLRRLQLSADPERPGRFRLEL
SEG      .XXXXXXXXXXXXXXXXXXXXXXXXXXXXX.....
PRD      cccccchhhhhhhhhhhhhhhhhhhhhccccccccchhhhhhhhhhhccccccccceeee
```

SEQ LGAGPGAVNLEWPLESVSYTIRGPTQHELQPPPGGPGTSLHFLNPQEAQRWAVLVRGAT
SEG
PRD cccccccceccccceeeeeccccccccccccccccccccceeeeeccccchhhhhheeeecce

```
SEQ      VEGQNGSKSNSPPALGPEACPVSLSPPPEASTLKGPPEADLPRSPGNLTEREELAGSLA
SEG      .....
PRD      eeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeechhhhhhhhhhhh
```

SEQ RAIAGGDEKGAAQVAVLAQHRVALSVQLQEACFPGP IRLQVTLEDAASAASAASHV
SEG XXXXXXXXXXXXX..
PRD hhhhcccchhhhhhhhhhhhhhhhhhccccccccccccceeeccchhhhhhhhhhhhee

SEQ ALQVHPHCTVAALQEQVFSELGFPPAVQRWVIGRLCVPERSLASYGVRQDGDPAFLYLL
SEG
PRD eeccccchhhhhhhhhhhhhccccchhhhhhhhhhhhhccccccccccccccccceeeec

```
SEQ  SAPREAPATGSPQHPQKMDGELGRLFPSPSLGLPPGPQPAASSLPSPLQPSWSCPSCFTI
SEG  .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX...
PRD  cccccccccchhhhhhhhhhhccccccccccccccccccccccccccccccccccccccc
```

```

SEQ      NAPDRPGCEMCSTQRPCTWDPLAAAST
SEG      .....
PRD      CCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

```

(No Prosite data available for DKFZphtes3 23n19.2)

(No Pfam data available for DKFZphtes3_23n19.2)

DKFZphtes3_26g22

group: intracellular transport/trafficking

DKFZphtes3_26g22 encodes a novel 898 amino acid protein with similarity to kinesins.

The novel protein contains a ATP/GTP-binding site motif A (P-loop) and a kinesin motor domain signature. Kinesin is a microtubule-associated force-producing protein that play a role in organelle transport. It is an oligomeric complex composed of two heavy chains and two light chains. The kinesin motor activity is directed toward the microtubule's plus end. The heavy chain contains a large globular N-terminal domain which is responsible for the motor activity of kinesin, which is known to hydrolyze ATP and to bind and move on microtubules. Several proteins involved in chromosome segregation and cell division contain this motor domain, such as *Drosophila* claret segregational protein (*ncd*), *Drosophila* kinesin-like protein (*nod*), human CENP-E and human mitotic kinesin-like protein-1 (MKLP-1). The novel protein is a new kinesin like proprotein.

The new protein can find application in modulating chromosome transport in mitosis and meiosis and modulation of cell division.

strong similarity to kinesins

Sequenced by EMBL

Locus: unknown

Insert length: 3032 bp

No poly A stretch found, no polyadenylation signal found

```
1 CTGAAGCGCT GGGAGGCGGA CATTAAAGTG AAGTGGTTGC GGTAACTTGG
51 CCTGGGCGCTG AAGTGAGTGA GAGGCACATG AAGAGAAGTA TTCAAGTATT
101 TATACAGATA GGAATCAAGA TAATCAACAA TGCTGTGCAC TGAGGAAGAC
151 CTGTGCCACC ATATGAAAGT AGTAGTTCGT GTACGTCGGG AAAACACATA
201 AGAAAAAGCA GCTGGATTTC ATAAAGTGGT TCATGTTGTG GATAAACATA
251 TCCTAGTTTT TGATCCCAAA CAAGAAGAAG TCAGTTTTTT CCATGGAAG
301 AAAACATACAA ATCAAAATGT TATAAGAAA CAAAATAAGG ATCTTAAATT
351 TGTATTGTAT GCTGTTTTTG ATGAAACGTC AACTCAGTCA GAAGTTTTTG
401 AACACACTAC TAAGCCAATT CTTCTAGTGT TTTTGAATGC ATATAATTGC
451 ACAGTACTTG CCTATGGTGC CACTGGTGCT GGGGAAGACC ACACTATGCT
501 AGGATCAGCT GATGAACCTG GAGTGATGTA TCTAACAAATG TTACACCTTT
551 ACAATGCAT GGATGAGATT AAAGAAGAGA AAATATGTAG TACTGCAGTT
601 TCATATCTGG AGGTATATAA TGAACAGATT CGTGATCTCT TAGTAAATTC
651 AGGGCCACTT GCTGTCCGGG AAGATACCCA AAAAGGGGTG GTCGTTTCATG
701 GACTTACTTT ACACCAGCCC AAATCCTCAG AAGAAATTTT ACATTTATTG
751 GATAATGGAA ACAAAAACAG GACACAACAT CCCACTGATA TGAATGCCAC
801 ATCTTCTCGT TCTCATGCTG TTTTCCAAAT TTACTTGCGA CAACAAGACA
851 AAACAGCAAG TATCAATCAA AATGTCCGTA TTGCCAAGAT GTCACTCATT
901 GACCTGGCAG GATCTGAGCG AGCAAGTACT TCCGGTGCTA AGGGGACCCG
951 ATTTGTAGAA GGCACAAATA TTAATAGATC ACTTTTAGCT CTTGGGAATG
1001 TCATCAATGC CTTAGCAGAT TCAAAGAGAA AGAATCAGCA TATCCCTTAC
1051 AGAAATAGTA AGCTTACTCG CTTGTTAAAG GATTCTCTTG GAGGAAACTG
1101 TCAAACATATA ATGATAGCTG CTGTTAGTCC TTCCTCTGTA TTCTACGATG
1151 ACACATATAA CACTCTTAAG TATGCTAACC GGGCAAAGGA CATTAAATCT
1201 TCTTTGAAGA GCAATGTTCT TAATGTCAAT AATCATATAA CTCAATATGT
1251 AAAGATCTGT AATGAGCAGA AGGCAGAGAT TTTATTGTTA AAAGAAAAAC
1301 TAAAGGCCTA TGAAGAACAG AAAGCCTTCA CTAATGAAAA TGACCAAGCA
1351 AAGTTAATGA TTTCAAACCC TCAGGAAAAA GAAATCGAAA GGTTCRAGA
1401 AATCCTGAAC TGCTTGTTCC AGAATCGAGA AGAAATTAGA CAAGAATATC
1451 TGAAGTTGGA AATGTTACTT AAAGAAAATG AACTTAAATC ATTCTACCAA
1501 CAACAGTGCC ATAAACAAAT AGAAATGATG TGTCTGAAG ACAAGTAGA
1551 AAAGGCCACT GGAACACGAG ATCATAGACT TGCAATGTTG AAAACTCGTC
1601 GCTCCTACCT GGAGAAAAGG AGGGAGGAGG AATTGAAGCA ATTTGATGAG
1651 AATACTAATT GGCTCCATCG TGTCGAAAAA GAAATGGGAC TCTTAAGTCA
1701 AAACGGTCAT ATTCCAAAGG AACTCAAGAA AGATCTTCAT TGTCACCATT
1751 TGCACCTCCA GAACAAAGAT TTGAAAGCAC AAATTAGACA TATGATGGAT
1801 CTAGCTTGTC TTCAGGAACA GCAACACAGG CAGACTGAAG CAGTATTGAA
1851 TGCTTTACTT CCAACCCATA GAAACAATA TTGCACATTA AAAGAAGCCG
1901 GCGTGTCAAA TGCTGCTTTT GAATCTGACT TCAAAGAGAT CGAACATTG
1951 GTAGAGAGGA AAAAAGTGGT AGTTTGGGCT GACCAAACTG CCGAACAACC
2001 AAAGCAAAAC GATCTACCAG GGATTTCTGT TCTTATGACC TTTCACAAC
2051 TTGGACCAGT TCAGCCTATT CCTTGTGCT CATCTTCAGG TGGAACTAAT
2101 CTGGTTAAGA TTCCTACAGA AAAAAGAACT CGGAGAAAAC TAATGCCATC
2151 TCCCTTGAAA GGACAGCATA CTCTAAAGTC TCCACCATTCT CAAAGTGTGC
2201 AGCTCAATGA TTCTCTTAGC AAAGAACTTC AGCCTATTGT ATATACACCA
2251 GAAGACTGTA GAAAAGCTTT TCAAATCCG TCTACAGTAA CCTTAATGAA
2301 ACCATCATCA TTTACTACAA GTTTTCAGGC TATCAGCTCA AACATAAACA
2351 GTGATAATTG TCTGAAAATG TTGTGTGAAG TAGCTATCCC TCATAATAGA
```

```

2401 AGAAAAGAAT GTGGACAGGA GGACTTGGAC TCTACATTTA CTATATGTGA
2451 AGACATCAAG AGCTCGAAGT GTAAATTACC CGAACAAGAA TCACTACCAA
2501 ATGATAACAA AGACATTTTA CAACGGCTTG ATCCTTCTTC ATTCTCAACT
2551 AAGCATTCTA TGCCTGTACC AAGCATGGTG CCATCTTACA TGGCAATGAC
2601 TACTGCTGCC AAAAGGAAAC GGAAATTAAC AAGTTCTACA TCAAACAGTT
2651 CGTTAACTGC AGACGTAAT TCTGGATTGG CCAAACGTGT TCGACAAGAT
2701 AATTCAAGTG AGAAGCACTT ACAAGAAAAC AAACCAACAA TGGAACATAA
2751 AAGAAACATC TGTAATAATA ATCCAAGCAT GGTTAGAAAA TTTGGAAGAA
2801 ATATTTCAAA AGGAAATCTA AGATAAATCA CTTCAAAACC AAGCAAAATG
2851 AAGTTGATCA AATCTGCTTT TCAAAGTTA TCAATACCCT TTCAAAAATA
2901 TATTTAAAT CTTTGAAAGA AGACCCATCT TAAAGCTAAG TTTACCCAAG
2951 TACTTTTCTG AAGCAGAAAA ATGAAACTCT TTGTTTCTT CTTTGTGTG
3001 CTAATAAAAT AAAATTTCAA AAGAAAAA AA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 130 bp to 2823 bp; peptide length: 898
 Category: strong similarity to known protein
 Classification: Cell structure/motility
 Prosite motifs: ATP_GTP_A (113-121)
 KINESIN_MOTOR_DOMAIN1 (252-264)

```

1 MSVTEEDLCH HMKVVVRVRP ENTKEKAAGF HKVVHVVDKH ILVFDPKQEE
51 VSFFHGGKKT NQNVIKKQNK DLKFVFDVAV DETSTQSEVF EHTTKPILRS
101 FLNGYNCTVL AYGATGAGKT HTMLGSADEP GVMYLTMLHL YKCMDEIKEE
151 KICSTAVSYL EVYNEQIRDL LVNSGPLAVR EDTQKGVVVH GLTLHQPKSS
201 EEILHLLDNG NKNRTQHPD MNATSSRSHA VFQIYLRQOD KTASINQNVN
251 IAKMSLIDLA GSERASTSGA KGTRFVEGTN INRSLALGN VINALADSKR
301 KNOHIPYRNS KLTRLLKDSL GGNCQTIMIA AVSPSSVFYD DTYNLTLYAN
351 RAKDIKSSSL SNVLNVNHHI TQYVKICNEQ KAEILLKKEK LKAYEEQKAF
401 TNENDQAKLM ISNPQEKEIE RFQEILNCLF QNREEIRQEY LKLEMLLKEN
451 ELKSFYQQQC HKQIEMMCSE DKVEKATGKR DHRLAMLKTR RSYLEKRREE
501 ELKQFDENTN WLRHVEKEMG LLSQNGHIPK ELKKDLHCHH LHLQNKDLKA
551 QIRHMDLAC LQEQQHRQTE AVLNALLPTL RKQYCTLKEA GLSNAAFESD
601 FKEIEHLVER KKVVVWADQT AEQPKQNDLP GISVLMTFPQ LGPVQPIPC
651 SSSGGTNLVR IPTEKRTRRK LMPSPKQGQ TLKSPSPSQSV QLNDLSLSEL
701 QPILVYTPEDC RKAFQNPSTV TLMKPSSFTT SFQAISNNIN SDNCLKMLCE
751 VAIPIHNRKE CGQEDLDSTF TICEIDIKSSK CKLPEQESLP NDNDKILQRL
801 DPSSFSTKHS MPVPSMVPSY MAMTTAAKRK RKLTSSTNS SLTADVNSGF
851 AKRVRQDNSS EKHLQENKPT MEHKNRICKI NPSMVRKFRG NISKGNLR

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_26g22, frame 1

SWISSPROT:YB3D_SCHPO PUTATIVE KINESIN-LIKE PROTEIN C2F12.13., N = 3,
 Score = 874, P = 9e-93

TREMBL:DMU89264_1 product: "kinesin like protein 67a"; Drosophila
 melanogaster kinesin like protein 67a mRNA, complete cds., N = 1, Score
 = 880, P = 4.2e-88

TREMBL:SPBC649_1 gene: "SPBC649.01c"; product: "putative kinesin-like
 protein"; S.pombe chromosome II cosmid c649., N = 3, Score = 814, P =
 9.8e-86

PIR:S64238 kinesin-related protein KIP3 - yeast (Saccharomyces
 cerevisiae), N = 2, Score = 802, P = 2.5e-83

>TREMBL:DMU89264_1 product: "kinesin like protein 67a"; Drosophila

melanogaster kinesin like protein 67a mRNA, complete cds.
Length = 814

HSPs:

Score = 880 (132.0 bits), Expect = 4.2e-88, P = 4.2e-88
Identities = 181/345 (52%), Positives = 238/345 (68%)

```
Query: 11 HMKVVVRVPENTKEAAGFHKKVHVVDKHLVFDPKQEEVSFF-HGKKTNTQNVIKKQN 69
      ++KV VRVRP N +E      ++ V+D+ L+FDP +E+ FF G K   +++ K+ N
Sbjct: 8 NIKVAVRVRPYNVRELEQKQRSIIKVMDSALLFDPDEEDDEFFQGAQPYRDITKRMN 67

Query: 70 KDLKFVFDVFDDETSTQSEVFHTTKPILRSFLNGYNCTVLAYGATGAGKTHTMLGSADE 129
      K L  FD VFD ++ ++FE T P++ + LNGYNC+V YGATGAGKT TMLGS
Sbjct: 68 KKLTFMEFDRVFDIDNSNQDLFEECTAPLVDAVLNGYNCVSFVYGATGAGKTFTMLGSEAH 127

Query: 130 PGVMYLTMLHLYKCMDEIKEEKICSTAVSYLEVYNEQIRDLLVNSGPLAVREDTQKGVVV 189
      PG+ YLTM L+ + + + VSYLEVYNE + +LL SGPL +RED GVVV
Sbjct: 128 PGLTYITMQDLFDKIQAQSDVRKFDVGVSYLEVYNEHVMNLLTKSGPLKLRDNN-GVVV 186

Query: 190 HGLTLHQPKSSEILHLLDNGNKNRTQHPTDMNATSSRSHAVFQIYLRQDKTASINQNV 249
      GL L  S+EE+L +L GN +RTQHPTD NA SSRSHA+FQ+++R ++ + V
Sbjct: 187 SGLCLTFPIYSAEELLRLMLGNSHRTQHPTDANAESSRSHAIFQVHIRITERKTDTKRTV 246

Query: 250 RIAKMSLIDLAGSERASTSGAKGTRFVEGTNINRSLALGNVINALADSKRKNQHIPPYRN 309
      K+S+IDLAGSERA+++ G RF EG +IN+SLLALGN IN LAD + HIPYR+
Sbjct: 247 ---KLSMIDLAGSERAASTKGIGVRFEGASINKSLALGNCINKLADGLK---HIPYRD 300

Query: 310 SKLTRLKDSLGGNCQTIMIAAVSPSSVFYDDTYNTLKYANRAKDI 355
      S LTR+LKDSLGGNC+T+M+A VS SS+ Y+DTYNTLKYA+RAK I
Sbjct: 301 SNLTRLKDSLGGNCRTLMVANVSMSSLTIEDTYNTLKYASRAKKI 346
```

Pedant information for DKFZphtes3_26g22, frame 1

Report for DKFZphtes3_26g22.1

```
[LENGTH] 898
[MW] 102281.63
[pI] 9.09
[HOMOL] SWISSPROT:YB3D SCHPO PUTATIVE KINESIN-LIKE PROTEIN C2F12.13. 3e-97
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YGL216w] 2e-88
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YGL216w] 2e-88
[FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YGL216w] 2e-88
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YGL216w] 2e-88
[FUNCAT] 09.10 nuclear biogenesis [S. cerevisiae, YPR141c] 5e-42
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YPR141c] 5e-42
[FUNCAT] 03.13 meiosis [S. cerevisiae, YPR141c] 5e-42
[FUNCAT] 11.01 stress response [S. cerevisiae, YPR141c] 5e-42
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins
[S. cerevisiae, YPR141c] 5e-42
[FUNCAT] 30.05 organization of centrosome [S. cerevisiae, YPR141c] 5e-42
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YKL079w]
4e-28
[BLOCKS] BL00411H
[BLOCKS] BL00411G
[BLOCKS] BL00411F
[BLOCKS] BL00411E Kinesin motor domain proteins
[BLOCKS] BL00411C Kinesin motor domain proteins
[BLOCKS] BL00411B Kinesin motor domain proteins
[BLOCKS] BL00411A Kinesin motor domain proteins
[SCOP] d2kin.1 3.29.1.5.3 Kinesin [Rat (Rattus norvegicus) 1e-117
[SCOP] d3kar_ 3.29.1.5.4 Kinesin [Baker's yeast (Saccharomyce 1e-112
[PIRKW] nucleus 6e-87
[PIRKW] heterodimer 4e-68
[PIRKW] DNA binding 9e-60
[PIRKW] heterotetramer 2e-54
[PIRKW] mitosis 9e-60
[PIRKW] microtubule binding 4e-68
[PIRKW] ATP 6e-87
[PIRKW] phosphoprotein 5e-59
[PIRKW] heterotrimer 4e-68
[PIRKW] purine nucleotide binding 1e-26
[PIRKW] P-loop 6e-87
[PIRKW] coiled coil 4e-68
[PIRKW] heptad repeat 3e-62
[PIRKW] methylated amino acid 2e-54
[PIRKW] hydrolase 2e-54
[PIRKW] GTP binding 1e-60
```

[PIRKW] cell division 5e-57
 [SUPFAM] kinesin-related protein KIP1 3e-50
 [SUPFAM] kinesin-related protein CIN8 7e-33
 [SUPFAM] kinesin heavy chain 2e-54
 [SUPFAM] suppressor protein SMY1 1e-26
 [SUPFAM] kinesin-related protein KIF3 4e-68
 [SUPFAM] kinesin-related protein KIF2 1e-46
 [SUPFAM] kinesin-related protein unc-104 7e-60
 [SUPFAM] unassigned kinesin-related proteins 6e-87
 [SUPFAM] centromere protein E 3e-54
 [SUPFAM] kinesin-related protein KLP61F 5e-57
 [SUPFAM] kinesin-related protein MKLP-1 2e-28
 [SUPFAM] pleckstrin repeat homology 7e-60
 [SUPFAM] kinesin-related protein KIF1B 4e-61
 [SUPFAM] kinesin motor domain homology 6e-87
 [SUPFAM] kinesin-related protein KLP4 1e-43
 [SUPFAM] kinesin-related protein nodA 1e-30
 [SUPFAM] kinesin-related protein Eg5 5e-59
 [PROSITE] ATP_GTP_A 1
 [PROSITE] KINESIN_MOTOR_DOMAIN1 1
 [PFAM] Kinesin motor domain
 [KW] Irregular
 [KW] 3D
 [KW] LOW_COMPLEXITY 8.57 %

SEQ MSVTEEDLCHHMKVVVRVPENTKEKAAGFHKKVVHVVDKHLVFDPKQEEVSFFHGKKT
 SEG
 3kar-TBEEE

SEQ NQNVIKKQNKDLKFVDAVFDETSTQSEVFEHTTKPILRSFLNGYNCTVLAYGATGAGKT
 SEG
 3kar- EEEETTTTTTEEEETEETTTTCHHHHHHHHHH-HHHGGGGCCCEEEETTTTCHH

SEQ HTMLGSADEPGVMYLTMLHLYKCMDEIKEEKICSTAVSYLEVYNEQIRDLLVNSGPLAVR
 SEG
 3kar- HHHHTTTT--THHHHHHHHHHHHHHHHHGGGCEEEEEEEETTEEEETT-TCCCEEE

SEQ EDTQKGVVHGLTLHQPKSSEELHLLDNGNKNRTQHPTDMNATSSRSHAVFQIYLRQQD
 SEG
 3kar- EETTTTEEEETTCCEEECCGGHHHHHHHHHHHCCTTTTCHHHHHHCCEEEEEEEEEE

SEQ KTASINQNVRIAKMSLIDLASERASTSGAKGTRFVEGTNINRSLALGNVINALADSKR
 SEG
 3kar- TTTTCEE---EEEEEECCCCCCCC---HHHHHHHHHHHHHHHHHHHHHHHTTTT

SEQ KNQHIPYRNSKLTRLKDSLGGNCQTIMIAAVSPSSVFYDDTYNTLKYANRAKDIKSSLK
 SEG
 3kar- TTTCTTTTTHHHHHHGGGCTTTTEEEEEECCCGGGHHHHHHHHHHHH.....

SEQ SNVLNVNHHITQYVKICNEQKAEILLKEKLKAYEEQKFTNENDQAKLMISNPQEKEIE
 SEG xxxxxxxx.....xxxxxxxxxxxxxxxxxxxxxx
 3kar-

SEQ RFQEILNCLFQNREEIRQEYLKLEMLKENELKSFYQQQCHKQIEMMCSEDKVEKATGKR
 SEG
 3kar-

SEQ DHRLAMLKTRRSYLEKRREEELKQFDENTNWLHRVEKEMGLLSQNGHIPKELKDLHCHH
 SEG
 3kar-

SEQ LHLQNKDLKAQIRHMDLACLQEQHRQTEAVLNALLPTLRKQYCTLKEAGLSNAAFESD
 SEG xxx.....
 3kar-

SEQ FKEIEHLVERKKVVWADQTAEQPKQNDLPGISVLMTFPQLGVPVQPIPCSSSSGGTNLVK
 SEG
 3kar-

SEQ IPTEKRTRRKLMPSPLKGQHTLKSPPSQSVQLNDSLSKELQPIVYTPEDCRKAFQNPSTV
 SEG
 3kar-

SEQ TLMKPSSFTTSFQAISSNINSDNCLKMLCEVAIPHNRKKECGQEDLDSTFTICEDIKSSK
 SEG
 3kar-

SEQ CKLPEQESLPNDNKDILQRLDPSSFSTKHSMPVPSMVPSYAMMTAAKRKRKLTSSSTNS
 SEG
 3kar-

SEQ SLTADVNSGFAKRVQRDNSSEKHLQENKPTMEHKRNICKINPSMVRKFGRNISKGNLR
 SEG xxx.....
 3kar-

Prosite for DKFZphtes3_26g22.1

PS00017 113->121 ATP_GTP_A PDOC00017
 PS00411 252->264 KINESIN_MOTOR_DOMAIN1 PDOC00343

Pfam for DKFZphtes3_26g22.1

HMM_NAME	Kinesin motor domain		
HMM	*RCRPLNeREindgcscvVQWPpWtGyktvnhghegds.....		
	R+RP N +E+++G +VV + + + + +E S		
Query	17 RVRPENTKEKAAGFHKVHVVD-KHILVFDPKQEEVSFFHGKKTNNQNV	64	
HMMphksFtFDHVFwWncTQedVYdtvAHPIVDDcFhGYNCTIFAYGQ		
	+ F+ED VF+ ++TQ +V++ + PI+ +++GYNCT++AYG		
Query	65 IKKQNKDLKFVFDVDETSTQSEVFETTKPILRSFLNGYNCTVLAYGA	114	
HMM	TGSGKTYTMMGpggehPDHmGIIPRcCHDIFdrIdkfgekDhdFwhVvKCS		
	TG+GKT+TM G + D+ G+ + +++++ D + + + +S		
Query	115 TGAGKTHTMLG----SADEPGVMYLTMLHLYKCMDEIK-EEKIC-STAVS	158	
HMM	YMEIYNEeIYDLLCPnPqhMkpLnIHEHPNMGPvYqGCTEfHvCSYeDac		
	Y+E+YNE+I+DLL+ N ++PL+++E+ G+ V G+T+ +S E+++		
Query	159 YLEVYNEQIRDLLV-N---SGPLAVREDTQKGVVHGLTLHQPKSSEEIL	204	
HMM	hWIWqGnknRHVAaTnMNdHSSRShtIFTIHVeQrHk..qcdehvcHSKM		
	H+++ GNKNR+ +T MN++SSRSH++F+I ++Q K + V++ KM		
Query	205 HLLDNGNKNRQTQHPDMDNATSSRSHAVFQIYLRQDKTASINQNVRIAKM	254	
HMM	NLVDLAGSERvnrTGAEGQRlKEGCNINqSLttLGnVinaLaDggTKYmY		
	+L+DLASER++ +GA G+R+ EG+NIN+SL++LGNVINALAD +		
Query	255 SLIDLASERASTSGAKGTRFVEGTNINRSLALGNVINALADSK-----	299	
HMM	gghgHIPYRDSKLTWLLQDSLGGNcKtCMIACIWPadWNYEETLSTLRYA		
	+++HIPYR SKLT+LL+DSLGGNC T MIA+++P+ + Y++T +TL+YA		
Query	300 RKNQHIPYRNSKLTRLKDSLGGNCQTIMIAAVSPSSVFYDDTYNTLKYA	349	
HMM	dRAKnIkNkPQINEDPcamalWRrYheQIdMKhgL*		
	+RAK+IK + N + + + +Y + + K++		
Query	350 NRAKDIKSSLKSNVLNVN-NHITQYVKICNEQKAEI	384	

DKFZphtes3_27d1

group: metabolism

DKFZphtes3_27d1 encodes a novel 712 amino acid protein similar to ubiquitin-specific proteases (EC 3.1.2.15).

The novel protein contains both, a ubiquitin carboxyl-terminal hydrolases family 2 signature 1 and signature 2. Pfam predicts a new member of the ubiquitin carboxyl-terminal hydrolases family 2. The ubiquitin system is responsible for the turn over of proteins. Ubiquitin carboxyl-terminal hydrolases (EC 3.1.2.15) (UCH) (deubiquitinating enzymes) are thiol proteases that recognize and hydrolyze the peptide bond at the C-terminal glycine of ubiquitin. These enzymes are involved in the processing of poly-ubiquitin precursors as well as that of ubiquitinated proteins.

The novel protein is a new member of the ubiquitin carboxyl-terminal hydrolases family 2, represented by proteins such as yeast UBP1-16, human tre-2, human isopeptidase T and others.

The novel protein can find application in modulation of ubiquitin- and protein metabolism in cells.

similarity to ubiquitin-specific proteases

complete cDNA, complete cds, 4 EST hits

Sequenced by GBF

Locus: unknown

Insert length: 2871 bp

Poly A stretch at pos. 2836, no polyadenylation signal found

```

1  CCAAACCTGA AAGAGGTTGA TTTGTAATGA TTTGCAGGGG GGCACCTGGAG
51  GCAGCGGCCA GGACTTTTCA CTTAGGAGAT CAGCATTTCG CCTGATGGAA
101 ACTGGGCGAT CCTGCAGGGA CTGACCTCTG AGTTATCCAA AGGCCGACCT
151 GGGGAAAGAC TGATTTTGTAG GTTTTAATAG TTTTCAGATG CTTCAAGTGT
201 TGTGAACAGA GACTTGTGTT GATTATGCAT TTCTCAGCTA GACTAAATAA
251 ATGCTAGCAA TGGATACGTG CAAACATGTT GGGCAGCTGC AGCTTGCTCA
301 AGACCATTCC AGCCTCAACC CTCAGAAATG GCACTGTGTG GACTGCAACA
351 CGACCGAGTC CATTGTTGGCT TGCCTTAGCT GCTCCCATGT TGCTGTGGGA
401 AGATATATTG AAGAGCATGC ACTCAAGCAC TTTCAAGAAA GCAGTCATCC
451 TGTTCGATTG GAGGTGAATG AGATGTACGT TTTTGTGTAC CTTTGTGATG
501 ATTATGTTCT GAATGATAAC GCAACTGGAG ACCTGAAGTT ACTACGACGT
551 ACATTAAGTG CCATCAAAAG TCAAAATTAT CACTGCACAA CTCGTAGTGG
601 GAGGTTTTTA CGGTCCATGG GTACAGGTGA TGATTCTTAT TTCTTACATG
651 ACGGTGCCCC ATCTCTGCTT CAAAGTGAAG ATCAACTGTA TACTGCTCTT
701 TGGCAGAGGA GAAGGATACT AATGGGTAAA ATCTTTTCGA CATGGTTTGA
751 ACAATCACCC ATTGGAAGAA AAAAGCAAGA AGAACCATTT CAGGAGAAAA
801 TAGTAGTAA AAGAGAAGTA AAGAAAAGAC GGCAGGAATT GGAGTATCAA
851 GTTAAAGCAG AATTGGAAAG TATGCCTCCA AGAAAAGATT TACGTTTACA
901 AGGGCTCGCT CAGTCGACCA TAATAGAAAT AGTTTCTGTT CAGGTGCCAG
951 CACAAACGCC AGCATCACCA GCAAAAGATA AAGTACTCTC TACCTCAGAA
1001 AATGAAATAT CTCAAAAAGT CAGTGACTCC TCAGTTAAAC GAAGGCCAAT
1051 AGTAACCTCT GGTGTAACAG GATTGAGAAA TTTGGGAAAT ACTTGCTATA
1101 TGAATTCTGT TCCTCAGGTG TTGAGTCATT TACTTATTTT TCGACAATGT
1151 TTTTAAAGC TTGATCTGAA CCAATGGCTG GCTATGACTG CTAGCGAGAA
1201 GACAAGATCT TGTAAGCATC CACCAGTCAC AGATACAGTA GTATATCAAA
1251 TGAATGAATG TCAGGAAAAA GATACAGGTT TTGTTTGCTC CAGACAATCA
1301 AGTCTGTCAT CAGGACTAAG TGGTGGAGCA TCAAAAGGTA GAAAGATGGA
1351 ACTTATTCAG CCAAAGGAGC CAACTTCACA GTACATTTC TTTTGTTCATG
1401 AATGTCATAC TTTGTTCCAA GTCATGTGGT CTGGAAGTGA GGCCTTGGTC
1451 TCACCATTG CTATGCTACA CTCAGTGTGG AGACTCATTC CTGCCTTTCG
1501 TGGTTACGCC CAACAAGACG CTCAGGAATT TCTTTGTGAA CTTTATAGATA
1551 AAATACAACG TGAATTAGAG ACAACTGGTA CCAGTTTACC AGCTCTTATC
1601 CCCACTTCTC AAAGGAAACT CATCAAAACA GTTCTGAATG TTGTAATAA
1651 CATTTTTCAT GGACAACCTC TTAGTCAGGT TACATGTCTT GCATGTGACA
1701 ACAAATCAAA TACCATAGAA CCTTCTGTTG ACTTGTCTAT GGAGTTTCCA
1751 GAAAGGTATC AATGCAGTGG AAAAGATATT GCTTCCAGC CATGCTCTGGT
1801 TACTGAAATG TTGGCCAAAT TTACAGAAAC TGAAGCTTTA GAAGGAAAAA
1851 TCTACGTATG TGACCAAGT AACTCAAAGC GTAGAAGGTT TTCCTCCAAA
1901 CCAGTTGTAC TCACAGAAGC CCAGAACAA CTTATGATAT GCCACCTACC
1951 TCAGGTCTCT AGACTGCACC TCAAACGATT CAGGTGGTCA GGACGTAATA
2001 ACCGAGAGAA GATTGGTGT CATGTTGGCT TTGAGGAAT CTTAAACATG
2051 GAGCCCTATT GCTGCAGGGA GACCCTGAAA TCCTCAGAC CAGAATGCTT
2101 TATCTATGAC TTGTCGCGCG TGGTGATGCA CCATGGGAAA GGATTTGGCT
2151 CAGGGCACTA CACTGCCTAC TGCTATAATT CTGAAGGAG GTTCTGGGTA
2201 CACTGCAATG ATTCCAACT AAGCATGTGC ACTATGGATG AAGTATGCAA
2251 GGCTCAAGCT TATATCTTGT TTTATACCCA ACGAGTTACT GAGAATGGAC

```

```

2301 ATTCTAAACT TTTGCCTCCA GAGCTCCTGT TGGGGAGCCA ACATCCCAAT
2351 GAAGACGCTG ATACCTCGTC TAATGAAATC CTTAGCTGAT CCAAAGACAA
2401 TGGGGTTTTT TTCCTGTGAT TTATATATAT ACTTTTAAAG AACTGATGT
2451 ACCATTTTAA ACTTCATTTT TTCTTGTGAA TCAGTGTATA CTACATTAT
2501 ACATTTTATA TCTAACAAAT TTTTTTTTTT ACAAAGTATA AATGTATATA
2551 TCAACTGAAG GTAACACTT TTTTCATATT TGGAGTTTTA AACTTTTGGT
2601 GTTTACCTCA GACTGATGTT ACCTCTTTTA TATTTTATG TCTTAATTGG
2651 CTCGGATGAT GAACCTGTGC AATCTTCTAC CAACAAAGT CAAGTGGCAT
2701 CATTTTATAT ACATGTATCT TTTTCAGGTA TTTTCTATAC AAATTCTTAA
2751 TAGATGGAAG ATTAGACTCT AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2801 AAAAAAAAAA AAAAAAAAAA AAGGGGCGGC CGCTCTAAAA AAAAAAAAAA
2851 AAAAAAAAAA AAAAAAAAAA G

```

BLAST Results

No BLAST result

Medline entries

98072201:
Regulation of ubiquitin-dependent processes by deubiquitinating enzymes.

98431658:
The ubiquitin system.

Peptide information for frame 2

ORF from 251 bp to 2386 bp; peptide length: 712
Category: similarity to known protein
Prosite motifs: UCH_2_1 (274-290)
UCH_2_2 (619-638)
UCH_2_2 (619-638)

```

1 MLAMDTCKHV GQLQLAQDHS SLNPQKWHCV DCNTTESIWA CLSCSHVACG
51 RYIEEHALKH FQESSHPVAL EVNEMYVFCY LCDDYVLNDN ATGDLKLLRR
101 TLSAIKSONY HCTTRSGRFL RSMGTGDDSY FLHDGAQSL QSEDQLYTAL
151 WHRRRILMGK IFRTWFEQSP IGRKKQEEPF QEKIVVKREV KRRRQEEYQ
201 VKAELESMP RPKSLRLQGLA QSTIIIEIVSV QVPAQTPASP AKDKVLSTSE
251 NEISQKVS DS SVKRRPIVTP GVTGLRNLGN TCYMSVVLQV LSHLLIFRQC
301 FLKLDLNLQW AMTASEKTRS CKHPPVTDTV VYQNECQEK DTGFVCSRQS
351 SLSSGLSGGA SKGRKMELIQ PKEPTSQYIS LCHELHTLFQ VMWSGKWALV
401 SFFAMLSHVW RLIPAFRGYA QQDAQEFLCE LLDKIQRELE TTGTSLPALI
451 PTSQRKLKIQ VLNVVNNIFH GQLLSQVTCL ACDNKSNTIE PFDLSLEFP
501 ERYQCSGKDI ASQPCLVTEM LAKFTETAL EGKIYVCDQC NSKRRRFSSK
551 PVVLTEAQKQ LMICHLPOVL RLHLKRFWS GRNNREKIGV HVGFEILNM
601 EPYCCRETLK SLRPECFIYD LSAVVMHGHK GFGSGHYTAY CYNSEGGFWV
651 HCNDKLSMC TMDEVCKAQA YILFYTORVT ENHSGKLLPP ELLLSQHPN
701 EDADTSSNEI LS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_27d1, frame 2

PIR:S57591 hypothetical protein YMR223w - yeast (Saccharomyces cerevisiae), N = 4, Score = 218, P = 8.4e-38

SWISSPROT:UBPB HUMAN UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 11 (EC 3.1.2.15) (UBIQUITIN THIOLESTERASE 11) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 13) (DEUBIQUITINATING ENZYME 11) (KIAA0055), N = 2, Score = 300, P = 9.3e-31

TREMBL:AF079565_1 gene: "Ubp41"; product: "ubiquitin-specific protease UBP41"; Mus musculus ubiquitin-specific protease UBP41 (Ubp41) mRNA, complete cds., N = 3, Score = 187, P = 8.7e-30

PIR:I58376 hypothetical protein unip - mouse, N = 3, Score = 214, P = 1.2e-28

>SWISSPROT:UBPB_HUMAN_UBIQUITIN_CARBOXYL-TERMINAL_HYDROLASE_11 (EC 3.1.2.15)
 (UBIQUITIN_THIOLESTERASE_11) (UBIQUITIN-SPECIFIC_PROCESSING_PROTEASE_13)
 (DEUBIQUITINATING_ENZYME_11) (KIAA0055).
 Length = 1,118

HSPs:

Score = 300 (45.0 bits), Expect = 9.3e-31, Sum P(2) = 9.3e-31
 Identities = 95/301 (31%), Positives = 149/301 (49%)

Query: 381 LCHELHTLFQVMWSGKVALVSPFAMLHVSRLIPAFRGYAQQDAQEFLCELLDKIQREL- 439
 + E + + +W+G++ +SP ++ ++ F GY+QD+QE L L+D + +L
 Sbjct: 826 VAEFEGIIKALWTGQYRISPKDFKITIGKINDQFAGYSQQDSQELLLFLMDGLHEDLN 885

Query: 440 -----ETTGTSLPALIPTSQRKLIKQVLN--VVNNIFHGQLLSQVTCCLADNKSNT 488
 E L + LN ++ +F GQ S V CL C KS T
 Sbjct: 886 KADNRKRYKEENDHLDLDFKAAEHAWQKHKLNESIIVALFQGQFKSTVQCLTCHKKSRT 945

Query: 489 IEPFWDLSLEFFERYQCSGKDIASQPCLVTEMLAKFTETEALEGKIYVCDQCNSKRRRFS 548
 E F LSL +C+ +D CL + +K E + + C C ++R
 Sbjct: 946 FEAFMYLSLPLASTSKCTLQD-----CL--RLFSK--EEKLTNNRIFYCSHCRARR---- 992

Query: 549 SKPVVLTEAQQLMICHLPQVLRHLKRFWRSGRNNREKIGVHVGF-EILNMEPYCC-- 605
 ++ K++ I LP VL +HLKRF + GR ++K+ V F E L++ Y
 Sbjct: 993 -----DSLKKIEIWKLPVLLVHLKRFSDGRW-KQKLQTSVDFPLENLDLSQYVIGP 1044

Query: 606 RETLKSRLPECFIYDLSAVVMHHGFGSGHYTAYCYNSEGGFWVHCNDSKLSMCTMDEV 665
 + LK Y+L +V H+G G GHYTAYC N+ W +D ++S ++ V
 Sbjct: 1045 KNNLKK-----YNLFSVSNHYG-GLDGGHYTAYCKNAARQRWFKFDHEVSDISVSSV 1096

Query: 666 CKAQAYILFYTQ---RVTE 681
 + AYILFYT RVT+
 Sbjct: 1097 KSSAAYILFYTSLGPRVTD 1115

Score = 126 (18.9 bits), Expect = 9.3e-31, Sum P(2) = 9.3e-31
 Identities = 41/116 (35%), Positives = 63/116 (54%)

Query: 200 QVKAELSMPPR--KSLRLQGLAQSTIIIEIVSVQVPAQTSPAPAKDKVLSTSENEISQKV 257
 Q+ AE + P + +S + Q+ I+ + P TP ++K + EIS ++
 Sbjct: 701 QIPAERDREPSKLKRSYSSPDITQA--IQEEKRKPTVTPTVNRENKPTCYPKAEIS-RL 757

Query: 258 SDSSVKR-RPIVT---PGVTGLRNLGNTCYMNSVLQVLS---HLLIF--RQCFLKLDLNQ 308
 S S ++ P+ P +TGLRNLGNTCYMNS+LQ L HL + R C+ D+N+
 Sbjct: 758 SASQIRNLNPVFGSGPALTGLRNLGNTCYMNSILQCLCNAPHLADYFNRCYQD-DINR 816

Score = 50 (7.5 bits), Expect = 8.3e-23, Sum P(2) = 8.3e-23
 Identities = 29/106 (27%), Positives = 51/106 (48%)

Query: 173 RKQEEPFQEKIVVKREVKKRRQELEYQVKAELSMPPRKSLRLQGLAQSTIIIEIVSVQV 232
 + KQE+ +E+ +++ K R++E E + K + E+ + Q A+ + + S Q
 Sbjct: 475 KNQEKELRERQEEQKEKLRKEEQKAKKKQEA-EENEITEKQQAKEEMKKESEQA 533

Query: 233 PAQ---TPASPAKD---KVLSTSENEIS--QKVSDDSSVKRRPIVTPGV 272
 + T A K+ K S SE+E S +K + KR P TP +
 Sbjct: 534 KKDCKETSAKRGKEITGVKRQSKSEHETSDAKKSVEDRGKRCP--TPEI 580

Score = 42 (6.3 bits), Expect = 5.7e-22, Sum P(2) = 5.7e-22
 Identities = 13/58 (22%), Positives = 27/58 (46%)

Query: 167 EQSPIGRKKQEEPFQEKIVVKREVKKRRQELEY-QVKAELSMPPRKSLRLQGLAQST 223
 EQ +KKQE E +++ K+ ++ E Q K E + ++ + G+ ++
 Sbjct: 498 EQEQKAKKKQEAENEITEKQQAKEEMKKESEQAKKEDKETSARKGKEITGVKRQS 555

Pedant information for DKFZphtes3_27d1, frame 2

Report for DKFZphtes3_27d1.2

[LENGTH] 712
 [MW] 81155.71
 [pI] 8.21
 [HOMOL] SWISSPROT:UBPB_HUMAN_UBIQUITIN_CARBOXYL-TERMINAL_HYDROLASE_11 (EC 3.1.2.15)
 (UBIQUITIN_THIOLESTERASE_11) (UBIQUITIN-SPECIFIC_PROCESSING_PROTEASE_13) (DEUBIQUITINATING
 ENZYME_11) (KIAA0055). 4e-32
 [FUNCAT] 06.13.01 cytoplasmic degradation [S. cerevisiae, YMR223w] 5e-33
 [FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation,
 palmitoylation, farnesylation and processing) [S. cerevisiae, YMR223w] 5e-33

```

[FUNCAT] 06.13 proteolysis [S. cerevisiae, YBL067c] 3e-19
[FUNCAT] 10.03.99 other osmosensing activities [S. cerevisiae, YDR069c] 2e-17
[FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YDR069c] 2e-17
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YDR069c] 2e-17
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDR069c] 2e-17
[FUNCAT] 09.25 vacuolar and lysosomal biogenesis [S. cerevisiae, YDR069c] 2e-17
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YNL186w] 4e-17
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YHL010c] 3e-12
[BLOCKS] BL00970A Nuclear transition protein 2 proteins
[BLOCKS] BL00972D
[BLOCKS] BL00972C
[BLOCKS] BL00972B
[BLOCKS] BL00972A
[EC] 3.1.2.15 Ubiquitin thiolesterase 5e-06
[PIRKW] alternative splicing 2e-11
[PIRKW] thiolester hydrolase 5e-06
[PIRKW] hydrolase 1e-14
[SUPFAM] RING finger homology 7e-11
[SUPFAM] deubiquinating enzyme SSV7 5e-16
[PROSITE] MYRISTYL 5
[PROSITE] AMIDATION 2
[PROSITE] CAMP_PHOSPHO_SITE 1
[PROSITE] CK2_PHOSPHO_SITE 10
[PROSITE] TYR_PHOSPHO_SITE 2
[PROSITE] UCH_2_2_1
[PROSITE] PKC_PHOSPHO_SITE 17
[PROSITE] ASN_GLYCOSYLATION 4
[PROSITE] UCH_2_1_1
[PFAM] Ubiquitin carboxyl-terminal hydrolases family 2
[PFAM] Ubiquitin carboxyl-terminal hydrolases family 2
[KW] Alpha Beta
[KW] LOW COMPLEXITY 4.92 %

```

```

SEQ      MLAMDTCCKHVGQLQLAQDHSSLNPQKWHCVDCNTTESIWACLSCHSVACGRYIEEHALKH
SEG      .....
PRD      cccccccccchhhhhhhccccccccccceccccceeeeeccccccccchhhhhhhhhhh

```

SEQ FQESSHPVALEVNEMYVFCYLCDYVLNDNATGDLKLLRRTLSAIKSQNYHCTTRSGRFL
 SEG
 PRD hhhhccceeeccccceeeeeccccccccccccchhhhhhhhhhhhhccceeecccccc

```
SEQ    RSMGTGDDSYFLHDGAQSLLQSEDQLYTALWHRRRILMGKIFRTWFEQSPIGRKKQEFPF
SEG    .....
PRD    cccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
```

```
SEQ      QEKIVVKREVKRRRQELEYQVKAELSMPPRKSLRLQGLAQSTIIIEIVSVQVPAQTFPASP
SEG      xxxxxxxxxxxxxxxxx.....
PRD      hhheehhhhhhhhhhhhhhhhhhhhhccccccccccccccceeeeeeccccccccccc
```

SEQ AKDKVLSTSENEISQKVDSSVKRRPIVTPGVTGLRNLGNTCYMNSVLQVLSHLLIFRCQ
SEG
PRD cchhhhhhhhhhhhhhhhhh

```
SEQ      FLKLDLNQWLAMTASEKTRSCKHPPVTDTVVYQMNECQEKDTGFVCSRQSSLSGLSGGA
SEG      .....XXXXXXXXXXXXXXX
PRD      hhhhhhchhhhhhhhhhhhhhhccccccceehhhhhccccccccccccccccccccccc
```

```
SEQ SKGRKMELIQPKETSQYISLCHELHTLFQVMWSGKWALVSPFAMLSVWRLIPAFRGYA
SEG      xxxxx.....
PRD      cccccceccccccccchhhhhhhhhhhhhhhhhhhhhccceeeccchhhhhhhhhhhcccccch
```

SEQ QQDAQEFLCELLDKIQRELETTGTSLPALIPTSQRKLIKQVLNVVNNIFHGQLLSQVTCL
SEG
PRD hhhhhhhhhhhhhhhhhhhhhhhccccccccchhhhhhhhhhhhhhhccccchhhhhhhhhc

```
SEQ      ACDNKSNTIEPFDLSLEFPERYQCSGKDIASQPCLVTEMLAKFTETEALLEGKIYVCDQC
SEG      .....
PRD      cccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhccccccc
```

SEQ NSKRRRFSSKPVVLTEAQQLMICHLPQVLRLHLKRFRWSGRNNREKIGVHVGFEIILNM
SEG
PRD ccccccccccchhhhhhhhhhhhhhhhchhhhhhhhhhhhhhhccccccccceeeeccccccc

SEQ EPYCCRET LKSLRPECFIYDL SAVVMHHGKGFSGHYTAYCYNSEGGFWVHCNDSKLSMC
 SEG
 PRD CC

```
SEQ      TMDEVCKAQAYILFYTQRVTEHGSKLLPPELLLGSHQPNEDADTSSNEILS
SEG      .....
PRD      cchhhhhhhhhhhhhhhheeecccccccccccccccccccccccccccccccccc
```

Prosites for DKFZphtes3_27d1.2

PS00001	33->37	ASN_GLYCOSYLATION	PDOC00001
PS00001	90->94	ASN_GLYCOSYLATION	PDOC00001
PS00001	484->488	ASN_GLYCOSYLATION	PDOC00001
PS00001	653->657	ASN_GLYCOSYLATION	PDOC00001
PS00004	545->549	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	6->9	PKC_PHOSPHO_SITE	PDOC00005
PS00005	113->116	PKC_PHOSPHO_SITE	PDOC00005
PS00005	116->119	PKC_PHOSPHO_SITE	PDOC00005
PS00005	213->216	PKC_PHOSPHO_SITE	PDOC00005
PS00005	254->257	PKC_PHOSPHO_SITE	PDOC00005
PS00005	261->264	PKC_PHOSPHO_SITE	PDOC00005
PS00005	315->318	PKC_PHOSPHO_SITE	PDOC00005
PS00005	320->323	PKC_PHOSPHO_SITE	PDOC00005
PS00005	394->397	PKC_PHOSPHO_SITE	PDOC00005
PS00005	453->456	PKC_PHOSPHO_SITE	PDOC00005
PS00005	506->509	PKC_PHOSPHO_SITE	PDOC00005
PS00005	542->545	PKC_PHOSPHO_SITE	PDOC00005
PS00005	548->551	PKC_PHOSPHO_SITE	PDOC00005
PS00005	580->583	PKC_PHOSPHO_SITE	PDOC00005
PS00005	608->611	PKC_PHOSPHO_SITE	PDOC00005
PS00005	611->614	PKC_PHOSPHO_SITE	PDOC00005
PS00005	676->679	PKC_PHOSPHO_SITE	PDOC00005
PS00006	125->129	CK2_PHOSPHO_SITE	PDOC00006
PS00006	164->168	CK2_PHOSPHO_SITE	PDOC00006
PS00006	223->227	CK2_PHOSPHO_SITE	PDOC00006
PS00006	247->251	CK2_PHOSPHO_SITE	PDOC00006
PS00006	249->253	CK2_PHOSPHO_SITE	PDOC00006
PS00006	313->317	CK2_PHOSPHO_SITE	PDOC00006
PS00006	506->510	CK2_PHOSPHO_SITE	PDOC00006
PS00006	525->529	CK2_PHOSPHO_SITE	PDOC00006
PS00006	661->665	CK2_PHOSPHO_SITE	PDOC00006
PS00006	706->710	CK2_PHOSPHO_SITE	PDOC00006
PS00007	193->200	TYR_PHOSPHO_SITE	PDOC00007
PS00007	192->200	TYR_PHOSPHO_SITE	PDOC00007
PS00008	218->224	MYRISTYL	PDOC00008
PS00008	355->361	MYRISTYL	PDOC00008
PS00008	359->365	MYRISTYL	PDOC00008
PS00008	471->477	MYRISTYL	PDOC00008
PS00008	589->595	MYRISTYL	PDOC00008
PS00009	171->175	AMIDATION	PDOC00009
PS00009	362->366	AMIDATION	PDOC00009
PS00972	274->290	UCH_2_1	PDOC00750
PS00973	619->638	UCH_2_2	PDOC00750

Pfam for DKFZphtes3_27d1.2

HMM_NAME	Ubiquitin carboxyl-terminal hydrolases family 2		
HMM	*GIQNLGNTCYMNSIIQCL*		
	G++NLGNTCYMNS++Q+L		
Query	274	GLRNLGNTCYMNSVLQVL	291
HMM_NAME	Ubiquitin carboxyl-terminal hydrolases family 2		
HMM	*YdLYgVICHYGntldyGHYWayVKNenhHRWkWYYFDDeTV*		
	YDL +V+ H+G + ++GHY+AY++N + ++W+ +D++		
Query	619	YDLSAVVMHHGKGFGSGHYTAYCYNSE--GGFWVHCNDSKL	657

DKFZphtes3_27k4

group: transmembrane protein

Summary DKFZphtes3_27k4 encodes a novel 490 amino acid protein with similarity to two hypothetical *C.elegans* proteins.

The novel protein contains 10 transmembrane regions and a leucine zipper. It is a member of the new 10 trans-membrane domain containing protein family which is specific for multicellular eukariotes.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes and as a new marker for testicular cells.

strong similarity to *C.elegans* K07H8.2/2K185.2
membrane regions: 10

complete cDNA, complete cds potential start at Bp 109, few EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1901 bp
Poly A stretch at pos. 1866, no polyadenylation signal found

```

1 GTGATTATACC AGAAAAACCA AGAAGACAGG CACAAAAAAG CAAACGGCAT
51 TTGGCAAGAT GGATTATCAA CTGCAGTACA GACTTTTAGT AATAGATCTG
101 AGCAACACAT GGAGTATCAC AGTTTCTCAG AGCAGTCTTT TCATGCCAAT
151 AATGGGCACG CATCATCAAG CTGCAGCCAA AAGTATGATG ACTATGCCAA
201 TTATAATTAC TGTGATGGAA GGGAGACTTC AGAAACCACT GCCATGTTAC
251 AAGATGAAGA TATATCTAGT GATGGTGATG AAGATGCTAT TGTAGAAGTG
301 ACCCCAAAAT TACCAAGGA ATCCAGTGGC ATCATGGCAT TGCAAATACT
351 TGTGCCCTTT TTGCTAGCTG GTTTTGGAAC AGTTTCAGCT GGCATGGTAC
401 TGGATATAGT ACAGCACTGG GAGGTGTTCA GAAAAGTTAC AGAAGTTTTC
451 ATTTTAGTCC CTGCACTTCT TGGTCTCAAA GGGAACTTGG AAATGACATT
501 GGCATCCAGA TTATCCACTG CAGTAAATAT TGGGAAGATG GATTCACCCA
551 TTGAAAAGTG GAACCTAATA ATTGGCAACT TGGCTTTAAA GCAGGTTTCA
601 GCAACAGTAG TGGGTTTCTC AGCAGCTGTG GCAGCAATTA TATTGGGCTG
651 GATTCCAGAA GGAATATATT ACCTTGATCA TTCCATACTT CTGTGCTCTA
701 GCAGTGTGGC AACTGCCTTC ATTGCATCTC TTCTGCAGGG AATAATAATG
751 TTGGGGTTA TCGTTGGTTC AAAGAAGACT GGTATAAATC CTGATAATGT
801 TGCTACACCC ATTGCTGCTA GTTTTGCGCA CCTTATAACT CTGCCCATAT
851 TGGCTTGGAT AAGTCAGGGC TTATACTCCT GTCTTGAGAC CTATTACTAC
901 ATTTCTCCAT TAGTTGGTGT ATTTTCTTGG GCTCTAACC CTATTGGAT
951 TATAATAGT GCCAAACATC CAGCCACAAG AACAGTTCTC CACTCAGGCT
1001 GGGAGCCTGT CATAACAGCT ATGGTTATAA GTAGCATTGG GGGCCTTATT
1051 CTGGACACAA CTGTATCAGA CCCAACTTGG GTTGGGATTG TTGTTTACAC
1101 GCCAGTTATT AATGGTATTG GTGGTAATTT GGTGGCCATT CAGGCTAGCA
1151 GGATTCTTAC CTACCTCCAT TTACATAGCA TTCCAGGAGA ATGTCCTGAT
1201 GAACCCAAAG GTTGTTACTA CCCATTTAGA ACTTCTTTG GTCCAGGAGT
1251 AAATAATAAG TCTGCTCAAG TTCTACTGCT TTTAGTGATT CCTGGACATT
1301 TAATTTCTCT CTACACTATT CATTTGATGA AAAGTGGTCA TACTTCTTTA
1351 ACTATAATCT TCATAGTAGT GTATTTATTT GCGCTGTGT TACAGGTATT
1401 TACCTTGCTG TGGATTGCTG ACTGGATGGT CCATCACTTC TGGAGGAAAG
1451 GAAAGGACCC GGATAGTTTC TCCATCCCCT ACCTAACAGC ATTGGGTGAT
1501 CTGCTCGGGA CAGCTCTGTT AGCCTTAAGT TTTCATTTTC TTGGCTTAT
1551 TGGAGATCGA GATGGAGATG TTGGAGACTA ATAAATTCTA CAAACTGCTC
1601 TCAAGTTACC AAGGAAGAAA ATACACGACA ACCACTTATG GCTCTTTTTC
1651 AAAACTCTTA AATCAGTAGT TTGACTTTTG CCAGGGTAAT CTTCAAGTTGG
1701 CCCTGATTCA ATTAATGGC CTTAATTTTT TTTAAGGAA TTTGTGTCAA
1751 AACCAGAATG AAGAGTATTC GTGCTGCTTT TCATAGAATA AATGATAATT
1801 TGACATAGAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
1851 AAAAAAAAAA AAGGGGAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAGGG
1901 G
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 109 bp to 1578 bp; peptide length: 490
Category: similarity to unknown protein

```

1 MEYHSFSEQS FHANNHGASS SCSQKYDDYA NYNYCDGRET SETTAMLQDE
51 DISSDGDEDA IVEVTPKLPK ESSGIMALQI LVPFLLAGFG TVSAGMVLDI
101 VQHWVEFRKV TEVFILVPAL LGLKGNLEMT LASRLSTAVN IGKMDSPIEK
151 WNLIIGNLAL KQVQATVVGFLAAVAAILG WIPEGKYLD HSILLCSSSV
201 ATAFIASLLQ GIIMVGIVG SKKTGINPDN VATPIAASFG DLITLAILAW
251 ISQGLYSCLE TYYYISPLVG VFFLALTPIW IIIAAKHPAT RTVLHSGWEP
301 VITAMVISSI GGLILDTTVS DPNLVGIVVY TPVINGIGGN LVAIQASRIS
351 TYLHLHSIPG ELPDEPKGCY YPFRTFFGPG VNNKSAQVLL LLVIPGHLIF
401 LYTIHLMKSG HTSLTIIFIV VYLFQAVLQV FTLLWIADWM VHHFWRKGD
451 PDSFSIPYLT ALGDLGLTAL LALSFLHLWL IGRDRGDVGD

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_27k4, frame 1

TREMBL:AF036704 2 gene: "ZK185.2"; Caenorhabditis elegans cosmid
ZK185., N = 1, Score = 730, P = 3.1e-72

TREMBL:AF047659 9 gene: "K07H8.2"; Caenorhabditis elegans cosmid
K07H8., N = 1, Score = 940, P = 1.7e-94

>TREMBL:AF047659 9 gene: "K07H8.2"; Caenorhabditis elegans cosmid K07H8.
Length = 507

HSPs:

Score = 940 (141.0 bits), Expect = 1.7e-94, P = 1.7e-94
Identities = 204/412 (49%), Positives = 271/412 (65%)

```

Query: 68 LPRESSGIMALQILVPFLLAGFGTVSAGMVLVDIVQHWVEFRKVEVFILVPALLGLKGNL 127
      +P ESS ++ Q+L PF +AG G V AG+VL IV W +F ++ E+ ILVPALLGLKGNL
Sbjct: 82 IPAESSYVLFQVLPFAVAGLGMVFAVLVSIVVTWPLFEEIPEILILVPALLGLKGNL 141

Query: 128 EMTLASRLSTAVNIGKMDSPIEKWNLIIGNLALKQVQATVVGFLAAVAAILGWIPEGKY 187
      EMTLASRLST N+G MDS ++ +++I NLAL QVQATVV FLA+ A L +IP G +
Sbjct: 142 EMTLASRLSTLANLGHMDSSKQRKDVVIANLALVQVQATVVAFLASAFAAALAFIPSGDF 201

Query: 188 YLDHSILLCSSSVATAFIASLLQGIIMVGIVGSKKTGINPDNVATPIAASFGDLITLAI 247
      H L+C+SS+ATA ASL+ ++MV VIV S+K INPDNVATPIAAS GDL TL +
Sbjct: 202 DWAHGLMCASSLATAACSASLVLSLLMVVIVITSRKYNINPDNVATPIAASLGDLTTTLTV 261

Query: 248 LAWISQGLYSCLETTYISPLVGVFLLALTPIWIIIAAKHPATRTVLHSGWEPVITAMVI 307
      LA+ T +++ +V V FL L P WI IA ++ T+ L++GW PVI +M+I
Sbjct: 262 LAFFGSVFLKAHNTESWLNVIIVIVLFLLLLPFWIKIANENEGTQETLYNGWTPVIMSMLI 321

Query: 308 SSIGGLILDTTVSDPNLVGIVVYTPVINGIGGNLVAIQASRISTYLHLHSIPGELPDEPK 367
      SS GG IL+T V + + Y PV+NG+GGNL A+QASR+STY H G LP+E
Sbjct: 322 SSAGGFILETAVRRYH--SLSTYGPVLNGVGGNLAQVQASRLSTYFHKAGTVGVLPNEWT 379

Query: 368 GCYYPF--RTFFGPGVNNKSAQVLLLVIPGHLIFLYTIHLM----KSGHTSLTIIFIVV 421
      + R FF +++SA+VLLLLV+PGH+ F + I L K+ T +F +
Sbjct: 380 VSRFTSVQRAFFSKEWDSRSARVLLLLVPGHICFNFLIQLFTLTSKNNVTPHGPLEFTSL 439

Query: 422 YLFQAVLQVFTLLWIADWMVHHFWRKGDVDPDSFSIPYLTALGDLGLTALLALSF 475
      Y+ A++QV LL++ +V W+ DPD+ IPYLTALGDLGLT LL + F
Sbjct: 440 YMIAAIQVVILLFVCQLLVALLWKWKIDPDNSVIPYLTALGDLGLTGLLFIVF 493

```

Pedant information for DKFZphtes3_27k4, frame 1

Report for DKFZphtes3_27k4.1

[LENGTH] 490
[MW] 53266.39

[pI] 5.29
 [HOMOL] TREMBL:AF047659_9 gene: "K07H8.2"; Caenorhabditis elegans cosmid K07H8. 4e-94

[PROSITE] LEUCINE ZIPPER 1
 [PROSITE] MYRISTYL 7
 [PROSITE] CAMP_PHOSPHO_SITE 1
 [PROSITE] CK2_PHOSPHO_SITE 7
 [PROSITE] PROKAR_LIPOPROTEIN 2
 [PROSITE] TYR_PHOSPHO_SITE 1
 [PROSITE] PKC_PHOSPHO_SITE 3
 [PROSITE] ASN_GLYCOSYLATION 1
 [KW] TRANSMEMBRANE 10
 [KW] LOW_COMPLEXITY 3.06 %

SEQ MEYHSFSEQSFHANNGHASSCSQKYDDYANYNYCDGRETSETTAMLQDEISSDGEDA
 SEG
 PRD ccc
 MEM

SEQ IVEVTPKLPKRESSGIMALQILVPFLLAGFGTVSAGMVLDIVQHWVEFRKVTEVFILVPAL
 SEG
 PRD eeeeecc
 MEMMM

SEQ LGLKGNLEMTLASRLSTAVNIGKMDSPIEKWNLIIGNLALKQVQATVVGFLAAVAAILG
 SEG
 PRD ccc
 MEM MMM

SEQ WIPEGKYLDHSILLCSSSVATAFIASLLQGIIMVGIVIGSKKTGINPDNVATPIAASFG
 SEG
 PRD hcc
 MEMMM

SEQ DLITLAILAWISQGLYSCLETYYYISPLVGVFLLALTPIWIIIAAKHPATRTVLHSGWEP
 SEG
 PRD ccc
 MEM MMM

SEQ VITAMVISSIGGLILDTTVSDPNLVGIVVYTPVINGIGGNLVAIQASRISTYLHLHSIPG
 SEG
 PRD hcc
 MEM MMM

SEQ ELPDEPKGCYYPFRTFFGPGVNNKSAQVLLLLVIPGHLIFLYTIHLMKSGHTSLTIIFIV
 SEG
 PRD ccc
 MEMMM

SEQ VYLFQAVLQVFTLLWIADWMVHFWRKGDPSFSIPYLTALGDLLGTALLALSFHFLWL
 SEGXXXXXXXXXXXXXXXXXXXXX.....
 PRD hcc
 MEM MMM

SEQ IGDRDGDVGD
 SEG
 PRD ecccccccccc
 MEM MM.....

Prosites for DKFZphtes3_27k4.1

PS00001	383->387	ASN_GLYCOSYLATION	PDOC00001
PS00004	108->112	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	23->26	PKC_PHOSPHO_SITE	PDOC00005
PS00005	65->68	PKC_PHOSPHO_SITE	PDOC00005
PS00005	221->224	PKC_PHOSPHO_SITE	PDOC00005
PS00006	5->9	CK2_PHOSPHO_SITE	PDOC00006
PS00006	54->58	CK2_PHOSPHO_SITE	PDOC00006
PS00006	146->150	CK2_PHOSPHO_SITE	PDOC00006
PS00006	238->242	CK2_PHOSPHO_SITE	PDOC00006
PS00006	257->261	CK2_PHOSPHO_SITE	PDOC00006
PS00006	296->300	CK2_PHOSPHO_SITE	PDOC00006
PS00006	318->322	CK2_PHOSPHO_SITE	PDOC00006
PS00007	25->33	TYR_PHOSPHO_SITE	PDOC00007
PS00008	90->96	MYRISTYL	PDOC00008
PS00008	122->128	MYRISTYL	PDOC00008
PS00008	216->222	MYRISTYL	PDOC00008
PS00008	220->226	MYRISTYL	PDOC00008

PS00008	254->260	MYRISTYL	PDOC00008
PS00008	336->342	MYRISTYL	PDOC00008
PS00008	339->345	MYRISTYL	PDOC00008
PS00013	12->23	PROKAR_LIPOPROTEIN	PDOC00013
PS00013	248->259	PROKAR_LIPOPROTEIN	PDOC00013
PS00029	459->481	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphtes3_27k4.1)

DKFZphtes3_27o14

group: testes derived

DKFZphtes3_27o14 encodes a novel 358 amino acid protein with similarity to *C. elegans* cosmid C55A6.

The new protein contains a C3HC4 zinc finger (RING finger) signature. The ring finger structure binds two atoms of zinc, and is involved in mediating protein-protein interactions. No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to *C. elegans* C55A6.1

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: /map="6"

Insert length: 2158 bp

Poly A stretch at pos. 2137, polyadenylation signal at pos. 2120

```

1 CCGAGGCCAG AGAGAAAAGA CTGCGAGGTG GCCGCAGCTG TGGCCGGAGA
51 GCACAAAGAA TGAACCAGCA GTGGAAGAGA AAATACTGTA AGCTGGCTGA
101 CTGCTGGTGA AGAAAATGCT TTATTTTGTG GGCAGGCATC TGTGGGATCT
151 GTAATAGAAA TATATTGGAG TAATTCAAGA TTCTGTGTTT GGCCCTTTTG
201 ACTGCTCTCT CTACAGGTTT AATTGGGCA TTTACTCATT TTCATGGCTC
251 CAAGGACCAT GTATGTGTTG GGGATCTTCA ATATTCATGT TATTTTCTCC
301 TTGGTCTTAT TATGATTGTT ACCTTTATGA AGCTTTAGTG ATTACAAAGC
351 ACTTTTCTTG TCCATTTTGA CCTGAGCTTT GTAACTCTG ATTTGCAGGA
401 TGGCTGGCTG TGGTGAAATT GATCATTCAA TAAACATGCT TCCTACAAAC
451 AGGAAAGCGA ACGAGTCCTG TTCTAATACT GCACCTTCTT TAACCGTCCC
501 TGAATGTGCC ATTTGTCTGC AAACATGTGT TCATCCAGTC AGTCTGCCCT
551 GTAAGCACGT TTTCTGCTAT CTATGTGTAA AAGGAGCTTC ATGGCTTGGG
601 AAGCGGTGTG CTCTTTGTCG ACAAGAAATT CCCGAGGATT TCCTTGACAA
651 GCCAACCTTG TTGTCACCAG AAGAAGCTCA GGCAGCAAGT AGAGGAAATG
701 GTGAATATGC ATGGTATTAT GAAGGAAGAA ATGGGTGGTG GCAGTACGAT
751 GAGCGCACTA GTAGAGAGCT GGAAGATGCT TTTCCAAAG GTAAAAAGAA
801 CACTGAAATG TTAATTGCTG GCTTTCTGTA TGTCGCTGAT CTTGAAAACA
851 TGGTTCAATA TAGGAGAAAT GAACATGGAC GTCGAGGAA GATTAAGCGA
901 GATATAATAG ATATACCAAA GAAGGGAGTA GCTGGACTTA GGCTAGACTG
951 TGATGCTAAT ACCGTAAACC TAGCAAGAGA GAGCTCTGCT GACGGAGCGG
1001 ACAGTGTATC AGCACAGAGT GGAGCTTCTG TTCAGCCCCC AGTGTCTTCT
1051 GTAAGGCCCC TAACATCAGT AGATGGTCAG TTAACAGGCC CTGCAACACC
1101 ATCCCCTGAT GCAAGCACTT CTCTGGAAGA CTCTTTGCTT CATTTACAAC
1151 TCAGTGGAGA CAACACAGCT GAAAGGAGTC ATAGGGGAGA AGGAGAAGAA
1201 GATCATGAAT CACCATCTTC AGGCAGGGTA CCAGCACCAG ACACCTCCAT
1251 TGAAGAAACT GAATCAGATG CCAGTAGTGA TAGTGAGGAT GTATCTGCAG
1301 TTGTTGCACA GCACTCCTTG ACCCAACAGA GACTTTTGGT TTCTAATGCA
1351 AACCAGACAG TACCCGATCG ATCAGATCGA TCGGGAAGTG ATCGATCAGT
1401 AGCAGGGGGT GGAACAGTGA GTGTCACTGT CAGATCTAGA AGGCCTGATG
1451 GACAGTGCAC AGTAACTGAA GTTTAAATAA AAATGCTTTC AGCTCCATGC
1501 TCAAGGTTGA AAGGGTTACC TGTAATTTTC TGCCACATA ACATTATACT
1551 CATCCCTAGT AGTGCATTTT GGGAGTTGGG GTGGGAAGGG GTATGGGAAG
1601 GATAGACTCA TAATTAATAA GTCTAACATG TCTCTGTGTA GAAATTTATT
1651 TAATGTAAGG AACTTGGGTG TTAATAGTTG AGAGCTGTTT AGTAATAACC
1701 CAGTTTCTTT GAGGTCTGTT TACTTTATAC TTTTAAAAA CTCTGTAGT
1751 TCTTTGGGCC AGTGTGTTTG TATTATCTGT GCATTAATGG TCCTCATCTG
1801 ACTCCTGCAT TGTGTCTTAT TTTTCTGCAT GGATTGGCAT AAGACCATAA
1851 CTAAAAATTT GCACCTGTGA GATGTTTGAT ATTATGAACA GGAACATAAA
1901 TTTAATGTAT GAATAGATGT GAATTTGGGA TTTCAAATAA GATGAATAAC
1951 AACTATTTTA TAGTAAAGTT ATTGAAATGG AAATGAAAC AGCCAGTAAC
2001 TTATGTTTCA GAATGTTTGT AACACACTTC ATGGTGTTC CATAGGCTTT
2051 GCTGTCTAGT CTTATAGTTT GAGGTTTTTT TGGTCTGCAT TTTTCTTTT
2101 GATTACAAAA TTTATAATTT AATAAATACT AGAGTTTATC AAAAAAAG
2151 AAAAAAAG

```

BLAST Results

Entry HSG117 from database EMBL:

human STS SHGC-36270.

Score = 1148, P = 8.9e-45, identities = 240/250

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 400 bp to 1473 bp; peptide length: 358
Category: similarity to unknown protein
Prosites motifs: ZINC_FINGER_C3HC4 (51-61)

```
1 MAGCGEIDHS INMLPTNRKA NESCSNTAPS LTVPECAICL QTCVHPVSLP
51 CKHVFCYLCV KGASWLGKRC ALCRQEIPED FLDKPTLLSP EELKAASRGN
101 GEYAWYYEGR NGWWQYDERT SRELEDAFSK GKKNTEMLIA GFLYVADLEN
151 MVQYRRNEHG RRRKIKRDII DIPKKGVAGL RLDCDANTVN LARESSADGA
201 DSVSAQSGAS VQPLVSSVRP LTSVDGQLTS PATPSPDAST SLEDSFAHLQ
251 LSGDNTAERS HRGEGEEDHE SPSSGRVPAP DTSIEETESD ASSDSEDVSA
301 VVAQHSLTQQ RLLVSNANQT VPDRSDRSRG DRSVAGGGTV SVSVRSRRPD
351 GQCTVTEV
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_27o14, frame 1

TREMBL:CEC55A6_1 gene: "C55A6.1"; Caenorhabditis elegans cosmid C55A6,
N = 2, Score = 165, P = 4.2e-15

SWISSPROT:YWZ6_CAEEL HYPOTHETICAL 39.3 KD PROTEIN C02B8.6 IN CHROMOSOME
X., N = 2, Score = 136, P = 3.1e-11

>TREMBL:CEC55A6_1 gene: "C55A6.1"; Caenorhabditis elegans cosmid C55A6
Length = 484

HSPs:

Score = 165 (24.8 bits), Expect = 4.2e-15, Sum P(2) = 4.2e-15
Identities = 42/106 (39%), Positives = 61/106 (57%)

```
Query: 75 QEIPEDFLDKPTLLSPEELKAASRGNGEYAWYYEGRN-GWWQYDERTSRELEDAFSKGGK 133
      Q +P LD ++ PEE K Y W Y G+N GWW+++ R RE+E+A++ GK
Sbjct: 93 QNVFALDLDA-SICDPEERK-----Y-WIYSGKNQGWWRFEPNREREIEEAYNAGKC 142
```

```
Query: 134 NTEMLIAGFLYVADLENMVQYRRNEHGRRRRIKR---DIID-IPKKGVAGL 180
      + E++I G YV D +QY R + R +KR D D I KG+AG+
Sbjct: 143 HCEVVICGRPYVIDFHQFLQYPRGVPNQARHVKRVSADDFDGIGVKGLAGI 193
```

Score = 96 (14.4 bits), Expect = 4.2e-15, Sum P(2) = 4.2e-15
Identities = 19/54 (35%), Positives = 30/54 (55%)

```
Query: 35 ECAICLQTCVHPVSLP-CKHVFCYLCVKGASW--LGKRCALCRQEIPEDFLDKPT 86
      EC IC + P ++P C H FC++C+KG +G C +CR I + +P+
Sbjct: 11 ECPICQCKMIVPTTIPACGHKFCFICLKGVMNDMGG-CPMCRGPIDSNIFAQPS 64
```

Pedant information for DKFZphtes3_27o14, frame 1

Report for DKFZphtes3_27o14.1

```
[LENGTH] 358
[MW] 38818.90
[pI] 5.17
[HOMOL] TREMBL:CEC55A6_1 gene: "C55A6.1"; Caenorhabditis elegans cosmid C55A6 2e-12

[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision
repair) [S. cerevisiae, YCR066w] 3e-04
[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YCR066w] 3e-04
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YCR066w] 3e-04
```

[FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation, palmitoylation, farnesylation and processing) [S. cerevisiae, YCR066w] 3e-04
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YDR265w] 4e-04
 [FUNCAT] 30.19 peroxisomal organization [S. cerevisiae, YDR265w] 4e-04
 [BLOCKS] BL00518 Zinc finger, C3HC4 type, proteins
 [PROSITE] MYRISTYL 2
 [PROSITE] AMIDATION 3
 [PROSITE] CAMP_PHOSPHO_SITE 1
 [PROSITE] CK2_PHOSPHO_SITE 12
 [PROSITE] TYR_PHOSPHO_SITE 1
 [PROSITE] ZINC_FINGER_C3HC4 1
 [PROSITE] PKC_PHOSPHO_SITE 9
 [PROSITE] ASN_GLYCOSYLATION 2
 [PFAM] Zinc finger, C3HC4 type (RING finger)
 [KW] Irregular
 [KW] 3D
 [KW] LOW_COMPLEXITY 19.83 %

SEQ MAGCGEIDHSINMLPTNRKANESCSNTAPSLTVPECAICLQTCVHPVSLPCKHVFCYLCV
 SEG
 lrmd-TTTTTEETTTEEETTEEEHHHH
 SEQ KGASWLKGRKALCRQEIPEDFLDKPTLLSPEELKAASRGNGEYAWYYEGRNGWWQYDERT
 SEG
 lrmd- HHHHHHCCBT TTTTCBCGGG-CBCC.....
 SEQ SRELEDAFSKGGKNTLEMIAGFLYVADLENMVQYRRNEHGRRRKIKRDIIDIPKKGVAGL
 SEGXXXXXXXXXXXXXXXXX.....
 lrmd-
 SEQ RLDCDANTVNLARESSADGADSVSAQSGASVQPLVSSVRPLTSVDGQLTSPATPSPDAST
 SEGXXXXXXXXXXXXX.....
 lrmd-
 SEQ SLEDSFAHLQLSGDNTAERSHRGEGEEDHESPSSGRVPAPDTSIEETESDASSDSEDVSA
 SEG x.....XXXXXXXXXXXXXXXXXXXXX.....
 lrmd-
 SEQ VVAQHSLSLQRLLVSNANQTVPDSDSGTDRSVAGGGTVSVSVRSRRPDGQCTVTEV
 SEG xxx.....XXXXXXXXXXXXXXXXXXXXX.....
 lrmd-

Prosites for DKFZphtes3_27o14.1

PS00001	21->25	ASN_GLYCOSYLATION	PDOC00001
PS00001	318->322	ASN_GLYCOSYLATION	PDOC00001
PS00004	132->136	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	16->19	PKC_PHOSPHO_SITE	PDOC00005
PS00005	120->123	PKC_PHOSPHO_SITE	PDOC00005
PS00005	217->220	PKC_PHOSPHO_SITE	PDOC00005
PS00005	260->263	PKC_PHOSPHO_SITE	PDOC00005
PS00005	274->277	PKC_PHOSPHO_SITE	PDOC00005
PS00005	325->328	PKC_PHOSPHO_SITE	PDOC00005
PS00005	330->333	PKC_PHOSPHO_SITE	PDOC00005
PS00005	343->346	PKC_PHOSPHO_SITE	PDOC00005
PS00005	346->349	PKC_PHOSPHO_SITE	PDOC00005
PS00006	32->36	CK2_PHOSPHO_SITE	PDOC00006
PS00006	89->93	CK2_PHOSPHO_SITE	PDOC00006
PS00006	120->124	CK2_PHOSPHO_SITE	PDOC00006
PS00006	195->199	CK2_PHOSPHO_SITE	PDOC00006
PS00006	222->226	CK2_PHOSPHO_SITE	PDOC00006
PS00006	240->244	CK2_PHOSPHO_SITE	PDOC00006
PS00006	282->286	CK2_PHOSPHO_SITE	PDOC00006
PS00006	287->291	CK2_PHOSPHO_SITE	PDOC00006
PS00006	293->297	CK2_PHOSPHO_SITE	PDOC00006
PS00006	320->324	CK2_PHOSPHO_SITE	PDOC00006
PS00006	328->332	CK2_PHOSPHO_SITE	PDOC00006
PS00006	354->358	CK2_PHOSPHO_SITE	PDOC00006
PS00007	98->107	TYR_PHOSPHO_SITE	PDOC00007
PS00008	329->335	MYRISTYL	PDOC00008
PS00008	337->343	MYRISTYL	PDOC00008
PS00009	66->70	AMIDATION	PDOC00009
PS00009	130->134	AMIDATION	PDOC00009
PS00009	159->163	AMIDATION	PDOC00009
PS00518	51->61	ZINC_FINGER_C3HC4	PDOC00449

Pfam for DKFZphtes3_27o14.1

```
HMM_NAME      Zinc finger, C3HC4 type (RING finger)
HMM            *CPICFCTFQlDyPWPFdePmMlPCgHsFCypCIrrW.....CPmC*
               C+IC      L      + P++LPC+H+FCY C++      C +C
Query          36 CAIC-----LQT---CVHPVSLPCKHVFCYLCVKGASWLGKRCALC      73
```

DKFZphtes3_28d14

group: testes derived

DKFZphtes3_28d14 encodes a novel 97 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1279 bp

Poly A stretch at pos. 1232, no polyadenylation signal found

```
1 GGAGCTCAGA AGTTGGGCAA AGGTCACAGC AGACTTCCTG AAAAGCAGAC
51 ACTGAGGAAC ACAGTGGAGA GCGGGAGTTC ACAGCGACGC AGCTGAGGAC
101 GACGCAGGAC CTCTCCCAAA GGTGCTGCAG CTCCAGCACC AGGGGCCAGG
151 GCTGCGGCGA CAGCAGCTCA GCAACCCCTG CTGTGCTCAA GTTCTTGCGG
201 ATTGAGAGCT AAGTTCAAAA TTTAGAAACA GTGCCTTAAA GACGGGCAAG
251 AAAACCCGGT GTGGGAGTCT GCTCATCTAT GGTGTGTAC TGCTCTCGCT
301 TTGATATTCT TAAATTCCCTA GGTACCAATG AAAAAGCCAA GTGAACGTGG
351 CAGAGTGAGG AGGAGACAGG AGCGTGTGCA CCTTCCATCT GTGAGAGGCA
401 CACTTCAGTC TGGGTTCAAG ATGCAGAATG GTGCCTACAG CAAAAAATAA
451 AAAAACACCC TCCTCCCTTC TTTACCATTT GAATGGACAT TTTCTTACC
501 TGTGATCCCA ACAGAAACAG ATCCAGACCT ATCATGTGAA GTCCACGTTC
551 CAGGATCAGA AGTAACCACT TTATGGACTG AGCTTACACG GGAAAGTCTA
601 CCCCCGACTC CTTCTGGATA GTACATACA CAGCTGCATA AAAACGTCTC
651 CAAGGGGACA TACGATGCAT TTGCTTGGTG TCCAGCCAA GCTCCCAACC
701 GCGGACCTCA CTGTTCTTGA GAGCTCGAGA GCTCGTCTCC TATCAATCAG
751 AGAACCCCAT CAGCTGTGAC CAACAGAGCT GGAGCCCTCT GTGGAGGGAG
801 CTGACCCAC ACACAGGACA GAGCAGAATC CTGATTATTT TACAAACTGC
851 AAACCTTCTG AGTAAGAAGA CAAAAATATA CATTCCAAGG TATCTGTAAA
901 GTGCTTGGA GATGCAGACA GCTGCACCGA GGGGCTCTGA TCCATCCACA
951 CGCTGCGGTT TGCTGCGGTC ACACACACGG TCTCAGTCAC GTGATGGTTT
1001 TGCTTTTATT TCTTAAACGG CTGAGTGATA ATCCAGCTAG TGTGCAGTCA
1051 TTTTCATACCT TTCAATGGGC GTCACCCGAG TGACGCTGCC CCAGCCCCAT
1101 GCTGAGGGCC GACACAATTC ACGGAACAGA TTCATCATAT TTGGTCTTTA
1151 TGTAAATAAT AAATGTTTTA AAATGCCTA AATATAAAAA AAAAAAAAAA
1201 AAAAAAAAAA AAAAAAAAAA AAAGGGCGGC CGAAAAAATA AAAAAAAAAA
1251 AAAAAAAAAA AAAAAAAAAA GGGCGGCCG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 328 bp to 618 bp; peptide length: 97
Category: putative protein

```
1 MKKPSEGRV RRRQERVHLP SVRGLQSGF KMONGAYSKK KKNTLLPSLP
51 FEWTFSLPVI PTETDPLDSC EVHVPGEVLT SLWTELTRTS LPPTPSG
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_28d14, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_28d14, frame 1

Report for DKFZphtes3_28d14.1

[LENGTH] 97
[MW] 10945.56
[pI] 9.80
[PROSITE] MYRISTYL 2
[PROSITE] CAMP_PHOSPHO_SITE 2
[PROSITE] CK2_PHOSPHO_SITE 2
[PROSITE] PKC_PHOSPHO_SITE 3
[KW] All_Alpha
[KW] LOW_COMPLEXITY 12.37 %

SEQ MKKPSEGRVRRRQERVHLPVSRGTLQSGFKMQNGAYSKKKKNTLLPSLPFEWTFSLPVI
SEGXX
PRD cccccchhhhhhhhhhhcc

SEQ PTETDPDLSCEVHVPGEVTSVLTSLWTELTRESLPPTPSG
SEG
PRD cccccccceeeccccccchhhhhhhhhhhcccccccc

Prosites for DKFZphtes3_28d14.1

PS00004	2->6	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	41->45	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	5->8	PKC_PHOSPHO_SITE	PDOC00005
PS00005	21->24	PKC_PHOSPHO_SITE	PDOC00005
PS00005	38->41	PKC_PHOSPHO_SITE	PDOC00005
PS00006	62->66	CK2_PHOSPHO_SITE	PDOC00006
PS00006	64->68	CK2_PHOSPHO_SITE	PDOC00006
PS00008	24->30	MYRISTYL	PDOC00008
PS00008	76->82	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_28d14.1)

DKFZphtes3_2a11

group: testes derived

DKFZphtes3_2a11 encodes a novel 1048 amino acid protein with very weak similarity to mucins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to mucin

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 4082 bp

Poly A stretch at pos. 4060, polyadenylation signal at pos. 4034

```
1 GAGGACTGCG AGCACAGCGG CGGCCGGGTG GCGGGGGTGA GTGGGGCCAG
51 CCGGGCTGGA CAGCAGCGGG CCCCAGGCGC CGCCGCCGCG ATCCCTCCCC
101 CGCGCCCGCG AGCACATCGC CGCCGCCGAG ATGGGCCCTC CGCGGCACCC
151 CCAGGCCGCG GAGATAGAAG CGGGCGGTGC GGGCGGCGGG CGCGGGCTAC
201 AGGTGGAAT GAGTTCTCAA CAGTTCTCTC GGTTAGGAGC CCCTCTACC
251 GGGCTGAGCC AGGCCCCCTC TCAGATTGCA AACAGTGGTT CTGCTGGATT
301 GATAAACCCA GCTGCTACAG TCAATGATGA ATCTGGTCGA GATTCTGAAG
351 TCAGTGCCAG GGAGCACATG AGTTCCAGCA GCTCCCTCCA GTCCCGGGAG
401 GAGAAGCAAG AGCCTGTTGT GGTAAAGGCC TATCCACAGG TGCAGATGTT
451 GTCGACACAC CATGCTGTCG CATCAGCCAC ACCTGTTGCA GTGACAGCCC
501 CGCCAGCACA CCTGACGCCA GCAGTGCCAC TTTTCATTTC GGAGGGACTT
551 ATGAAGCCGC CCCGGAAGCC CACCATGCCT AGCCGTCCCA TTGCTCCTGC
601 TCCACCTTCT ACCCTGTCAC TTCCCCCAA GGTTCAGGG CAGGTTACCG
651 TTACCATGGA GAGTAGCATC CCTCAAGCTT CAGCCATTCC TGTGGCAACA
701 ATCAGTGGAC AACAGGGCCA TCCAGTAAC CTGCATCACA TCATGACTAC
751 AAATGTGCAA ATGTCTATCA TCCGCAGCAA TGCTCCTGGG CCCCCTCTTC
801 ACATTGGAGC TTCTCATTTA CCTCGAGGTG CAGCTGTGTC TGCTGTGATG
851 TCCAGTTCTA AAGTAACCAC AGTCTGAGG CCGACCTCAC AGCTGCCAAA
901 TGCTGCTACT GCTCAGCCAG CAGTACAGCA CATCATTAC CAACCAATCC
951 AGTCTCGGCC ACCTGTGACC ACCTCCAATG CCATCCCTCC TGCTGTGGTA
1001 GCAACTGTCT CAGCCACCAG AGCTCAGTCT CCAGTCATCA CTACGACAGC
1051 GGGCATGCT ACTGATTGAG CACTTAGTAG GCCAACCTTG TCTATCCAGC
1101 ATCCTCCATC TGCAGCAATC AGTATTGAGC GTCCTGCCCA GTCACGAGAT
1151 GTACACAACA GAATCACACT ACCATCTCAC CCTGCATTAG GGACGCCAAA
1201 ACAGCAGCTT CATACAATGG CTCAGAAAAC AATCTTCAGT ACTGGCAGCG
1251 CAGTGGCTGC AGCCACAGTA GCACCTATTT TGGCAACCAA CACCATTCCT
1301 TCAGCGACCA CAGCTGGATC TGTGTACAC ACGCAAGCTC CCACAAGTAC
1351 CATGTGTACC ATGACAGTAC CCTCCCATTG CTCCCATGCT ACTGCTGTGA
1401 CCACCTCAAA CATCCCAGTC GCCAAGGTGG TGCCCCAGCA GATCACGCAC
1451 ACTTCTCCTC GGATCCAGCC AGACTACCCT GCCGAGAGGA GTAGCCTGAT
1501 TCCCATCTCC GGACATCGGG CCTCTCCCAA TCCTGTGGCC ATGGAAACCC
1551 GAAGTGACAA CAGACCGTCT GTTCCCGTTC AGTTCCAATA TTTTGTGCCA
1601 ACTTACCCCT CTCTGCATA CCCACTGGCG GCACATACCT ACACCCCAAT
1651 CACCAAGTCC GTGTCCACTA TCCGACAGTA TCCAGTTTCA GCTCAGGCTC
1701 CAAACTCTGC CATCACAGCT CAGACTGGTG TTGGGGTAGC GTCTACCGTC
1751 CACCTAAACC CCATGCAGTT GATGACAGTG GATGCATCGC ATGCTCGACA
1801 TATTTCAAGG ATCCAGCCAG CACCCATCAG TACCCAGGGT ATCCAGCCGG
1851 CCCCCATTGG GACCCCAGGG ATACAGCCTG CACCACTTGG CACACAGGGA
1901 ATTCACTCAG CAACCCCAAT CAACACACAA GGGCTTCAGC CTGCACCTAT
1951 GGGTACTCAG CAGCCTCAGC CTGAAGGAAA GACTTCAGCA GTGGTGTGG
2001 CAGATGGAGC CACAATTGTG GCCAACCCTA TTAGCAATCC ATTCAGTGCT
2051 GCTCCAGCAG CAACAACCGT GGTGCAAGCC CACAGCCAGA GTGCTAGCAC
2101 CAACGCTCCC GCCCAGGGCT CATCGCCACG GCCAAGCATA CTCCGGAAGA
2151 AACCTGCCAC AGATGGTGCC AAACCCAAGT CTGAAATCCA CGTGTCTATG
2201 GCCACTCCGG TCACTGTGTC CATGGAGACT GTATCCAATC AAAATAATGA
2251 TCAGCCTACC ATTGCCGTCC CTCCAAGTGC CCAGCAGCCC CCACCGACCA
2301 TTCCAATAT GATTGCAGCA GCCAGTCCCC CGTCACAACC AGCCGTTGCC
2351 CTTTCAACCA TTCTGGAGC GGTCCCCATC ACTCCACCCA TCACCACCAT
2401 TGCAGCTGCA CCACCTCCAT CAGTCACTGT GGGTGGCAGT CTTTCCTCCG
2451 TCTTGGGCCC TCCCGTTCCT GAAATTAAG TGAAGAAGA AGTAGAACCA
2501 ATGGATATCA TGAGGCCAGT TTCTGCAGTT CCTCCACTGG CTACCAACAC
2551 TGTGTCTCCA TCTCTGCAT TGCTGGCAAA CAACTTGTCC ATGCCATCAA
2601 GTGACCTACC ACCTGGTGCC TCCCAAGGA AAAAGCCTCG AAAGCAACAG
2651 CATGTGATCT CAACAGAAGA AGGTGACATG ATGGAGACAA ACAGCACTGA
2701 TGATGAGAA TCCACTGCCA AGAGTCTTCT GGTGAAGGCT GAGAAGCGCA
```

```

2751 AGTCTCCTCC CAAGGAGTAT ATTGATGAGG AAGGTGTGAG ATATGTCCCA
2801 GTGCGTCCAA GACCCCCCAT TACTTTGCTT CGTCACTATC GGAAACCCCTG
2851 GAAAGCTGCT TACCACCACT TTCAGAGGTA CAGTGACGTC CGGGTCAAAG
2901 AGGAGAAGAA AGCTATGCTG CAGGAAATAG CTAATCAGAA AGGAGTATCC
2951 TGTCGTGCTC AAGGCTGGAA AGTCCACCTC TGTGCTGCCC AGTTACTACA
3001 GCTGACGAAT CTAGAACATG ATGTCTATGA AAGACTTACT AACCTGCAGG
3051 AAGGGATTAT CCCAAAGAAA AAAGCAGCAA CAGATGATGA TCTCCACCGA
3101 ATAAACGAAC TGATACAGGG AAATATGCAG AGGTGTAAAC TTGTGATGGA
3151 TCAAATCAGT GAAGCCAGAG ACTCCATGCT TAAGGTTTTA GATCATAAAG
3201 ACCGTGCTCT GAAGCTGCTT AACAAAGACG GGACTGTCAA AAAAGTGTCC
3251 AAATTGAAGC GAAAGGAAAA AGTCTAGACC CAGAACAATC AGGAGATTGG
3301 AAGCAAATTT ATGAAGAAATG ATGGTGGGGG TGGGGGGAGG GTTTTGGTTT
3351 TTTCCAAAGT GGAACATTGA AATAAAGGAA GTGTTCTTGA GTTCCCGTGT
3401 GAAAGCAGAG GAACCCATGA CATCCAAGGG CGTGAAGGA TCAGAGCTGA
3451 CTGGACATAG TGAGCTGCCT TCTTGCCTTC GGGTGACACC CTGTTAAACC
3501 TGATCTGTGT CATAAGTGAC TCCGGATGCA TCAGTGTCCA CCAGTTGGAA
3551 GCAATGACAA GGATGGCTGG CTGGTGT TTTT TCAGCTTCC GGTTTATAGA
3601 CTGTATTTAT CTAGTGGATT CCTGCAGGCC CCATACTGAG CCTGGACTGA
3651 AAGTATCCAC TCGGACCATC TGTATCTCT CTACACTGAA AATAAAACCT
3701 CTTCACCCA CCCCATTCGG TTCTTCTGCC TGACCTTCAA ATGCCCATGT
3751 TGGCCTTTTA CAGCAGTGCC ACGGCACCAA GCGAGCTGCC ACATCTCACA
3801 CTCTAAAGGG TTGAACATAT TAGTTCTTGT CATTTTTTAA AAAAAACCAT
3851 TCCCAAGTGA AATTGTTATA TCGTCTGTCT TCGGTGTGTC AGAACTGGGT
3901 TTTTGTGGAG GTTCAGAGCA GGCAACACCA TAAGTTGCTC TCAGATCCTT
3951 GTTCTGAAGT ACATTCTTGG TTATCTGTAC TTCTGTAGCT GGTGTGATGC
4001 TGTTAATTGT ATGTACCACA CATCTCCAGA CGTTAATAAA GGACTCAAAG
4051 AGGTTTTTGT AAAAAAAAAA AAAAAAAAAA AA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 131 bp to 3274 bp; peptide length: 1048
 Category: similarity to known protein

```

1 MGPPRHPQAG EIEAGGAGGG RRLOVEMSSQ QFPRLGAPST GLSOAPSQIA
51 NSGSAGLINP AATVNDESGR DSEVSAREHM SSSSSLSQRE EKQEPVVVRP
101 YPQVQMLSTH HAVASATPVA VTAPPAHLTP AVPLSFSEGL MKPPPKPTMP
151 SRPIAPAPPS TSLPLPKVPG QVTVTMESSI PQASAIPVAT ISGQQGHPSN
201 LHHIMTTNQG MSIIRSNAPG PPLHIGASHL PRGAAAAAVM SSKVTTVLR
251 PTSQLPNAAT AQPAVQHIIH QPIQSRPPVT TSNAIPPAV ATVSATRAQS
301 PVITTTAAHA TDSALSRPTL SIQHPPSAAI SIQRPAQSRD VTTRITLPSH
351 PALGTFKQQL HTMAQKTIFS TGTPVAAATV APILATNTIP SATTAGSVSH
401 TQAPTSTIIV MTVPSSHSHA TAVTTSNIPV AKVVPQQITH TSPRIQPDYP
451 AERSSLIPIS GHRASPNPVA METRSDNRPS VPVQFYFLP TYPPSAYPLA
501 AHTYTPITSS VSTIRQYPVS AQAPNSAITA QTGVGVASTV HLNPMQLMTV
551 DASHARHIQG IQPAPISTQG IQPAPIGTPG IQPAPLGTQG IHSATPINTQ
601 GLQPAPMGTO QPQPEGKTSV VVLADGATIV ANPISNPFSA APAATTVVQT
651 HSQSASTNAP AQGSSPRPSI LRKKPATDGA KPKSEIHVSM ATPVTVSMET
701 VSNQNDQPT IAVPPTAQQP PPTIPTMIAA ASPPSQPAVA LSTIPGAVPI
751 TPPITTIAAA PPPSVTVGGS LSSVLGPPVP EIKVKEEVPE MDIMRPVSAV
801 PPLATNTVSP SLALLANNLS MPTSDLPFGA SPRKKPRKQQ HVIESTEEDM
851 METNSTDDEK STAKSLLVKA EKRKSPPKEY IDEEGVRYVP VRPRPPIITL
901 RHYRNPWKAA YHHFQRYSDV RVKEEKKAML QEIANQKQVS CRAQGWKVHL
951 CAAQLLQLTN LEHDVYERLT NLQEGIIPKK KAATDDDLHR INELIQGNMQ
1001 RCKLVMDQIS EARDSMLKVL DHKDRVLKLL NKNGTVKKVS KLKRKEKV

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2all, frame 2

SWISSPROT:MUC2_HUMAN MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2)., N = 1,
 Score = 334, P = 2.4e-25

PIR:A43932 mucin 2 precursor, intestinal - human (fragments), N = 1,
Score = 321, P = 3.2e-24

TREMBL:D88440_1 product: "high molecular mass nuclear antigen"; Gallus
gallus mRNA for high molecular mass nuclear antigen, partial cds., N =
1, Score = 312, P = 8.3e-24

PIR:S48478 glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast
(Saccharomyces cerevisiae), N = 1, Score = 300, P = 2.1e-22

>SWISSPROT:MUC2_HUMAN MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).
Length = 5,179

HSPs:

Score = 334 (50.1 bits), Expect = 2.4e-25, P = 2.4e-25
Identities = 184/770 (23%), Positives = 263/770 (34%)

Query: 96 VVVRYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
V P P T + + T V T P TP + + P P PT P
Sbjct: 3471 VTPTPTPTGTQTPTTTTPIITTTTPTPTPTGTQTPTTTTPIITTTTPTPTPTGTQTPT 3530

Query: 155 A-PAPPSTLSLPPKVP-GQVTVMESIPQASAI PVATISGQQGHPSNLHHIMTINVQMS 212
P +T P P G T T + P T +G Q P+ TT V +
Sbjct: 3531 TTPITTTTPTPTPTGTQTPTTTTPIITTTTPTPTPTGTQT-PTTPTITTTTPTPT 3589

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTVLRPTSQLPNAATAQPAVQHI 268
+ P P+ + P +++ +TT T T P I
Sbjct: 3590 PPTGTQTPTTTTPIITTTTPTPTPTGTQTPTTTTPIITTTTPTPTPTGTQTPTTTPI 3649

Query: 269 IHQPIQSRPPVTTSSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328
+ P T P T + T +P T T T + T++ P
Sbjct: 3650 TTTTPTPTPTPTGTQTPTTTTPIITTTTPTPTPTGTQTPTTTTPIITTTTPTPTPT 3706

Query: 329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
Q P + TT P+ GT + T + T TP T PI
Sbjct: 3707 PTGTQTPTTTTPIITTTTPTPTPTGTQTPTTTTPIITTTTPTPTPTGTQTPTTTPIIT 3766

Query: 386 TNTI-PSATTAGSVSHQTAPSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQIHTHTSP 443
T T+ P+ T G+ + T P +T T+T P+ + T TT V P T T
Sbjct: 3767 TTTTPTPTPTGTQTPTTTTPIITTTTPTPTPTGTQTPTTTTPIITTTTPTPTPTPTGT 3825

Query: 444 RIQPDYPAERSSSLIPISGHRASPNPVAMETRSNRPVSVQFYFL-PTYPPSAYPLAAH 502
+ P ++ + +P P +T + + P+ + PT P+
Sbjct: 3826 QTPTTTTPIITTTTPTPTPTPTGTQTPTTTTPIITTTTPTPTPTPTGT--TQTP 3874

Query: 503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNFMQLMTVDASHARHIQG 560
T TPIT++ + T P Q P + IT T V T Q T
Sbjct: 3875 TTTPTTTTPTPTPTGTQTPTTTTPIITTTTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 3932

Query: 561 IQPAPISTQGIQAPIGTPTGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTOQQPQ- 613
P P TQ PI T P P GTQ + TPI T P P GTQ P
Sbjct: 3933 PPTPTGTQTPTTTTPIITTTTPTPTPTGTQT-TPTTTTPIITTTTPTPTPTGTQTPTT 3991

Query: 614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAGSSPRPSIL 671
P T+ V T P + P + T T T +Q+ +T ++ P+
Sbjct: 3992 TPITTTTPTPTPTGTQTPTTTTPIITTTTPTPTPTGTQTPTTTTPIITTTTPTPTPT 4051

Query: 672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNNDOPTIAPV---PTAQPPPTIPTMI 728
T P + TP +T + T P PT Q P T P
Sbjct: 4052 PTGTQTPTTTTPIITTTTPTPTPTGTQTPTTTTPIITTTTPTPTPTGTQTPTTTPIIT 4111

Query: 729 AAASPPSQPAVALSTIPGAVPITPPITTTIAAAPPSPS-----VTVGSLSSVLGP-PVPEI 782
P+ T P PIT TT+ P P+ T + ++ + P P P
Sbjct: 4112 TTTTPTPTPTGTQT-PTTPTITTT-TTPTPTPTGTQTPTTTTPIITTTTPTPTPTPTG 4169

Query: 783 KVKEEVEPMDIMRPVSAVP-PLATNTVSPSLALLANNLSMPTSDLPFGASPRKKPRKQOH 841
P+ V+ P P T T P+ A + TS+ PP +S + R
Sbjct: 4170 TQTPTTTTPIITTTTPTPTPTGTGTGPPHTSTAPIAELTTSNPPPESTPQTSRSTSS 4229

Query: 842 VISTEEGDMMET 853
+ TE ++ T
Sbjct: 4230 PL-TESTTLLST 4240

Score = 328 (49.2 bits), Expect = 1.0e-24, P = 1.0e-24
Identities = 180/745 (24%), Positives = 254/745 (34%)

Query: 96 VVVRYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
V P P T + + T V T P TP + + P P PT P
Sbjct: 3540 VTPTPTPTGTQTPTTTTPIITTTTPTPTPTGTQTPTTTTPIITTTTPTPTPTGTQTPT 3599

Query: 155 A-PAPPSTLSLPPKVP-GQVVTMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212
P +T P P G T T + P T +G Q P+ TT V +
Sbjct: 3600 TTPITTTT VTPPTPTGTQTPTTPTITTTT VTPPTPTGTQT-PTTTPITTTT VTPPT 3658

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTTVLRETSQLPNAATAQPAVQHI 268
+ P P+ + P +++ +TT T T P I
Sbjct: 3659 PTPGTQTPTTTPITTTT VTPPTPTGTQTPTTTPITTTT VTPPTPTGTQTPTTTP I 3718

Query: 269 IHQPIQSRPFVTT SNAIPPAVVATVSATRAQSPVITTTAAHATDSALS RPTLSIQHPPSA 328
+ P T P T + T +P T T T + T++ P
Sbjct: 3719 TTTT VTPPTPTGTQTPTTTPITTTT VTPPTPTGTQTPTTTPITTTT VTP---PTPT 3775

Query: 329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTFVAAAT--VAPILA 385
Q P + TT P+ GT + T + T TP T PI
Sbjct: 3776 PTGTQTPTTTPITTTT VTPPTPTGTQTPTTTPITTTT VTPPTPTGTQTPTTTPITT 3835

Query: 386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTT SNIPVAKVVPQQITH TSP 443
T T+ P+ T G+ + T P +T T+T P+ + T TT V P T T
Sbjct: 3836 TTTVTPPTPTGTQTPTTTPITTTT VTPPTPTGTQTPTTTPITTTT VTPPT-TPGT 3894

Query: 444 RIQPDYPAERSSLIPISGHRASPNVAMETRSDNRPSVPVQFYFL-PTYPPSAYPLAAH 502
+ P ++ + +P P +T + + P+ + PT P+
Sbjct: 3895 QTPTTTPITTTT VTP-----PTPTPTGTQTPT---TTPITTTT VTPPTPTG--TQTP 3943

Query: 503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQ 560
T TPIT++ + T P Q P + IT T V T Q T
Sbjct: 3944 TTTPTTTT VTPPTPTGTQTPTTTPITTTT VTPPTP--TGTQTPTTTPITTTT VTP 4001

Query: 561 IQPAPISTQGIQAPIGTPTGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTOQFQ- 613
P P TQ PI T P P GTQ + TPI T P P GTQ P
Sbjct: 4002 PTPPTGTQTPTTTPITTTT VTPPTPTGTQ-TPTTTPITTTT VTPPTPTGTQTPTT 4060

Query: 614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGS SPRSIL 671
P T+ V T P + P + T T T +Q+ +T ++ P+
Sbjct: 4061 TPITTTT VTPPTPTGTQTPTTTPITTTT VTPPTPTGTQTPTTTPITTTT VTPPT 4120

Query: 672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNNQDPTIAVP---PTAQPPPTIPTMI 728
T P + TP +T + T P PT Q P T P
Sbjct: 4121 PTGTQTPTTTPITTTT VTPPTPTGTQTPTTTPITTTT VTPPTPTGTQTPTTTPITT 4180

Query: 729 AAASPPSQPAVALSTIPGAVPITPPITIAAA-PPPSVTVGGLSSVLGPPVPEIKVKEE 787
P+ T P T PI + + PPP + + S P +
Sbjct: 4181 TTTVTPPTPTGTQTGPPTHSTAPIAELTTSNPPFESSTPQTSRSTSSPLTESTTLLST 4240

Query: 788 VEPMDIMRPVSAVPPLATNTVSPSLALLANNLSMP--TSDLPPGASPR 833
+ P M S PP +T T +P+ + LS P T+ PPG R
Sbjct: 4241 LPPAEM--TSTAPP-STPT-APTTTSGGHTLSPPPSTTTSPPGTPTR 4284

Score = 325 (48.8 bits), Expect = 2.2e-24, P = 2.2e-24
Identities = 186/782 (23%), Positives = 261/782 (33%)

Query: 96 VVVRYPYQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
V P P T + + T V T P TP + + P P PT P
Sbjct: 3494 VTPPTPTGTQTPTTTPITTTT VTPPTPTGTQTPTTTPITTTT VTPPTPTGTQTPT 3553

Query: 155 A-PAPPSTLSLPPKVP-GQVVTMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212
P +T P P G T T + P T +G Q P+ TT V +
Sbjct: 3554 TTPITTTT VTPPTPTGTQTPTTTPITTTT VTPPTPTGTQT-PTTTPITTTT VTPPT 3612

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTTVLRETSQLPNAATAQPAVQHI 268
+ P P+ + P +++ +TT T T P I
Sbjct: 3613 PTPGTQTPTTTPITTTT VTPPTPTGTQTPTTTPITTTT VTPPTPTGTQTPTTTP I 3672

Query: 269 IHQPIQSRPFVTT SNAIPPAVVATVSATRAQSPVITTTAAHATDSALS RPTLSIQHPPSA 328
+ P T P T + T +P T T T + T++ P
Sbjct: 3673 TTTT VTPPTPTGTQTPTTTPITTTT VTPPTPTGTQTPTTTPITTTT VTP---PTPT 3729

Query: 329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTFVAAAT--VAPILA 385
Q P + TT P+ GT + T + T TP T PI
Sbjct: 3730 PTGTQTPTTTPITTTT VTPPTPTGTQTPTTTPITTTT VTPPTPTGTQTPTTTPITT 3789

Query: 386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTT SNIPVAKVVPQQITH TSP 443
T T+ P+ T G+ + T P +T T+T P+ + T TT V P T T
Sbjct: 3790 TTTVTPPTPTGTQTPTTTPITTTT VTPPTPTGTQTPTTTPITTTT VTPPT-TPGT 3848

Query: 444 RIQPDYPAERSSLIPISGHRASPNVAMETRSDNRPSVPVQFYFL-PTYPPSAYPLAAH 502
+ P ++ + +P P +T + + P+ + PT P+
Sbjct: 3849 QTPTTTPITTTT VTP-----PTPTPTGTQTPT---TTPITTTT VTPPTPTG--TQTP 3897

Query: 503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQ 560

T TPIT++ + T P Q P + IT TV T Q T
Sbjct: 3898 TTTTPIITTTTPTPTPTGTQTPTTTPIITTTTPTPTPT--TGTQTPTTPIITTTTPT 3955

Query: 561 IQPAPISTQGIQAPIGTPI---QPAPLGTQGIHSATPINTQGL---QPAPMGTOQPQ- 613
P P TQ PI T P P GTQ + TPI T P P GTQ P
Sbjct: 3956 PTPTPTGTQTPTTPTITTTTPTPTPTGTQ-TPTTPIITTTTPTPTPTGTQTPTT 4014

Query: 614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQSSPRPSIL 671
P T+ V T P + P + T T T +Q+ +T ++ P+
Sbjct: 4015 TPITTTTPTPTPTGTQTPTTPTITTTTPTPTPTGTQTPTTPTITTTTPTPTPT 4074

Query: 672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNNDOPTIAVP---PTAQPPPTIPTMI 728
T P + TP +T + T P PT Q P T P
Sbjct: 4075 PTGTQTPTTPTITTTTPTPTPTGTQTPTTPTITTTTPTPTPTGTQTPTTPTT 4134

Query: 729 AAASPPSQPAVALSTIPGAVPITPITTTIAAAPPSPVTVGGSLSVLPVPEIKVKEEV 788
P+ T P PIT TT P P+ T G+ + P I V
Sbjct: 4135 TTTVPTPTPTGTQT-PTTPTIT---TTTTVPTPTPT--GTQT---PTTPTITTTT 4184

Query: 789 EPMDIMRPVSAVPLATNTVSPSLALLANNLSMPTSDLPFGASPRKKPRKQHVISTEEG 848
P PP T+T +P L +N P S P + P + + +
Sbjct: 4185 TPTPTPTGTQTGPPTHTST-APIAELTTSN-PPESSTPQTSRSTSSPLTESTLLSTLP 4242

Query: 849 DMMETNSTDDEKSTAKSLLVKAERKSP 877
+E ST + SPP
Sbjct: 4243 PAIEMTSTAPPSTPTAPTTSGGHTLSPP 4271

Score = 324 (48.6 bits), Expect = 2.8e-24, P = 2.8e-24
Identities = 170/717 (23%), Positives = 248/717 (34%)

Query: 95 PVVVRYPQVQMLSTHHAVASATP--VAVTAPPAHLTPAVPLSFSEGLMKPPPKPTMPSR 152
P P P +T +P T PP TP+ P++ + + P P+ P
Sbjct: 1401 PPTTTPSPPTTTTTLPTTTPSPPTTTTTPPTTTPSPPTTTTTPPL-PTTTPSPPI 1459

Query: 153 PIAPAPPSTLSLPPKVPQVTVTMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212
PP+T PP T S + P T + P I +
Sbjct: 1460 TTTTTPPTTTPSPPTTTPSPPTTTPSPPTTTPPTTTPSPPTTTPSPPTTTPSPPTTTP 1516

Query: 213 IIRSNAPGPHLGHASHLPRGAAAAVMSKKVTVLRPTSQ--LPNAATAQPAVQHIIH 270
+ + P PP + P S T + PTS LP T P
Sbjct: 1517 LPPTTTPSPPTTTTTPPP-----TTTSPPTTTPITPPTSTTTLPTTTPSPPTTTT 1571

Query: 271 QPIQSRP-PVTTSNAPPAVVATVSA-TRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328
P + P P TT+ PP + T T SP TTT + S PT + PP++
Sbjct: 1572 PPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1631

Query: 329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKTI FSTGTPVAAATVAPILATNT 388
++ T T P P TP T I +T TP T + + T
Sbjct: 1632 TTTLPPTTTPSPPTTTTTP--PPTTTPSPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTT 1689

Query: 389 IPSATTAGSVSHTOAPTSTIVTMTVPSSHSHATAV-TTSNIPVAKVVPQIHTHSPRIQ 447
P TT + S T P+S I T T PS ++ + TT P P T T + P
Sbjct: 1690 SPPTTMTTPSPPTTTPSSPTTTTTPSSTTTPSPPTTMTTPSPPTTTPSPPTTMTTLP 1749

Query: 448 DYPAERSSLIPISGHRASPNPAMETRSNRPSPVPV-QFOYFLPTYPPSAY-P-----LA 500
+ + P+ P T + P VP+ + +L + P+ + P L
Sbjct: 1750 TTTSSPLTTTLPSPITPTTFSPFSTTTPTPCVPLCNWTGWLDSGKPNFHKPGGDELI 1809

Query: 501 AHTYTFITSSVSTIR--QYP-VSAQAPNSAITAQTVGV-VASTVHLNPMQLMTVDASHAR 556
P ++ + R YP V + VG + P ++ + A
Sbjct: 1810 GDVCGPGWAANISCRATMYPDVPVQLGQTVVCDVSLICKNEDQKPGGVIPM-AFCLN 1868

Query: 557 HIQGIQAPAPISTQGIQAPIGTPIQI-PAPLGTQGIHSATPINTQGLQAPMGTOQPQ-- 613
+ +Q TQ P + T + P P T I + T + P P GTQ P
Sbjct: 1869 YEINVQCCECVTQ---PTTMTTTTENPTPTTPTITTTTPTPT---PTPTGTQTPTT 1922

Query: 614 PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQSSPRPSILR 672
P T+ V T P + P + T T T +Q+ +T ++ P+
Sbjct: 1923 PITTTTPTPTPTGTQTPTTPTITTTTPTPTPTGTQTPTTPTITTTTPTPTPT 1982

Query: 673 KKPATDGAKPKSEIHVSMATPVTVSMETVSNQNNDOPTIAVP---PTAQPPPTIPTMIA 729
T P + TP +T + T P PT Q P T P
Sbjct: 1983 TGTQTPTTPTITTTTPTPTPTGTQTPTTPTITTTTPTPTPTGTQTPTTPTT 2042

Query: 730 AASPPSQPAVALSTIPGAVPITPITTTIAAAPPSPVTVGGSLSVLPVPEIKVKEEVE 789
P+ T P PIT TT P P+ T G+ + P V
Sbjct: 2043 TTVTPTPTPTGTQT-PTTPTIT---TTTTVPTPTPT--GTQTPTTPTITTTTPTPTPT 2096

Query: 790 PMDIMRPVSAVPLATNTVSPS 811
P P + P T TV+P+
Sbjct: 2097 PTGTQTPTTPTITTTTPTPTPT 2117

Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23
Identities = 174/717 (24%), Positives = 243/717 (33%)

Query: 96 VVVRPYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
V P P T + + T V T P TP + + P P PT P
Sbjct: 2068 VTPTPTPTGTQTPTTTPITTTTIVTPTPTPTGTQTPTTTPITTTTIVTPTPTPTGTQTPT 2127

Query: 155 A-PAPPSTLSLPPKVP-GQVTVMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212
P +T P P G T T + P T +G Q P+ TT V +
Sbjct: 2128 TTPITTTTIVTPTPTPTGTQTPTTTPITTTTIVTPTPTPTGTQT-PTTTPITTTTIVTPT 2186

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTTVLRPTSQLPNAATAQPAVQHI 268
+ P P+ + P +++ +TT T T P I
Sbjct: 2187 PTPGTQTPTTTPITTTTIVTPTPTPTGTQTPTTTPITTTTIVTPTPTPTGTQTPTTTP 2246

Query: 269 IHQPIQSRPPVTTSSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328
+ P T P T + T +P T T T + T++ P
Sbjct: 2247 TTTTIVTPTPTPTGTQTPTTTPITTTTIVTPTPTPTGTQTPTTTPITTTTIVT---PTPT 2303

Query: 329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
Q P + TT P+ GT + T + T TP T PI
Sbjct: 2304 PTGTQTPTTTPITTTTIVTPTPTPTGTQTPTTTPITTTTIVTPTPTPTGTQTPTTTPITT 2363

Query: 386 TNIT-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQITHTSP 443
T T+ P+ T G+ + T P +T T+T P+ + T TT V P T T
Sbjct: 2364 TTTVTPPTPTPTGTQTPTTTPITTTTIVTPTPTPTGTQTPTTTPITTTTIVTPTPTPTGT 2422

Query: 444 RIQPDYPAERSSLIPISGHRASPNPVAMETRSNDRPSVPVQFQYFL-PTYPPSAYPLAAH 502
+ P ++ + +P P +T + + P+ + PT P+
Sbjct: 2423 QTPTTTPITTTTIVT-----PTPTPTGTQTPT---TTPITTTTIVTPTPTPTG--TQTP 2471

Query: 503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQ 560
T TPIT++ + T P Q P + IT T V T Q T
Sbjct: 2472 TTPITTTTIVTPTPTPTGTQTPTTTPITTTTIVTPTPTPT--TGTQTPTTTPITTTTIVT 2529

Query: 561 IQPAPISTQGIQAPIGTPGI---QPAPLGTQGIHSATPINTOGL---QPAPMGTOQPO- 613
P P TQ PI T P P GTQ + TPI T P P GTQ P
Sbjct: 2530 PTPPTGTQTPTTTPITTTTIVTPTPTPTGTQ-TPTTTPITTTTIVTPTPTPTGTQTPTT 2588

Query: 614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAAT-TVVQTHSQSASTNAPAQGSSPRPSIL 671
P T+ V T P + P + T T T +Q+ +T ++ P+
Sbjct: 2589 TPITTTTIVTPTPTPTGTQTPTTTPITTTTIVTPTPTPTGTQTPTTTPITTTTIVTPTPT 2648

Query: 672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNDQPTIAPV---PTAQPPPTIPTMI 728
T P + TP +T + T P PT Q P T P
Sbjct: 2649 PTGTQTPTTTPITTTTIVTPTPTPTGTQTPTTTPITTTTIVTPTPTPTGTQTPTTTPITT 2708

Query: 729 AAASPPSQPAVALSTIPGAVPITPPITTTIAAPPPSVTVGSSLSSVLGPPVPEIKVKEEV 788
P+ T P PIT TT P P+ T G+ + P V
Sbjct: 2709 TTTVTPPTPTPTGTQT-PTTTPIT---TTTIVTPTPTPT--GTQTPTTTPITTTTIVTPTPT 2762

Query: 789 EPMDIMRPVSAVPPLATNTVSPS 811
P P + P T TV+P+
Sbjct: 2763 TPTGTQTPTTTPITTTTIVTPTPT 2784

Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23
Identities = 174/717 (24%), Positives = 243/717 (33%)

Query: 96 VVVRPYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
V P P T + + T V T P TP + + P P PT P
Sbjct: 2206 VTPTPTPTGTQTPTTTPITTTTIVTPTPTPTGTQTPTTTPITTTTIVTPTPTPTGTQTPT 2265

Query: 155 A-PAPPSTLSLPPKVP-GQVTVMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212
P +T P P G T T + P T +G Q P+ TT V +
Sbjct: 2266 TTPITTTTIVTPTPTPTGTQTPTTTPITTTTIVTPTPTPTGTQT-PTTTPITTTTIVTPTPT 2324

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTTVLRPTSQLPNAATAQPAVQHI 268
+ P P+ + P +++ +TT T T P I
Sbjct: 2325 PTPGTQTPTTTPITTTTIVTPTPTPTGTQTPTTTPITTTTIVTPTPTPTGTQTPTTTP 2384

Query: 269 IHQPIQSRPPVTTSSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328
+ P T P T + T +P T T T + T++ P
Sbjct: 2385 TTTTIVTPTPTPTGTQTPTTTPITTTTIVTPTPTPTGTQTPTTTPITTTTIVT---PTPT 2441

Query: 329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
Q P + TT P+ GT + T + T TP T PI
Sbjct: 2442 PTGTQTPTTTPITTTTIVTPTPTPTGTQTPTTTPITTTTIVTPTPTPTGTQTPTTTPITT 2501

Query: 386 TNIT-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQITHTSP 443
T T+ P+ T G+ + T P +T T+T P+ + T TT V P T T

Sbjct: 2502 TTTVTPTPTGTQTPTTPTITTTTTVTPTPTGTQTPTTPTITTTTTVTPTPTPTGT 2560

Query: 444 RIQPDYPAERSSLIPISGHRASPNPVMETRSDNRPSVPVQFYFL-PTYPPSAYPLAAH 502
 + P ++ + +P P +T + + P+ + PT P+

Sbjct: 2561 QTPTTPTITTTTTVT-----PTPTPTGTQTPT----TTPITTTTTVTPTPTPTG--TQTP 2609

Query: 503 TYTPITSSVS-TIRQYVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQG 560
 T TPIT++ + T P Q P + IT T V T Q T

Sbjct: 2610 TTTPTITTTTTVTPTPTPTGTQTPTTPTITTTTTVTPTPTPT--TGTQTPTTPTITTTTTVT 2667

Query: 561 IQPAPISTQGIQAPIGTPTGI---QPAPLGTQGIHSATPINTOGL---QPAPMGTOQQP- 613
 P P TQ PI T P P GTQ + TPI T P P GTQ P

Sbjct: 2668 PTPTPTGTQTPTTPTITTTTTVTPTPTGTQ-TPTTPTITTTTTVTPTPTPTGTQTPTT 2726

Query: 614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSIL 671
 P T+ V T P + P + T T T +Q+ +T ++ P+

Sbjct: 2727 TPITTTTTVTPTPTPTGTQTPTTPTITTTTTVTPTPTPTGTQTPTTPTITTTTTVTPTPT 2786

Query: 672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNNDQPTIAVP---PTAQPPPTIPTMI 728
 T P + TP +T + T P PT Q P T P

Sbjct: 2787 PTGTQTPTTPTITTTTTVTPTPTPTGTQTPTTPTITTTTTVTPTPTPTGTQTPTTPTITT 2846

Query: 729 AAASPPSQPAVALSTIPGAVPITPPITTTAAAPPSVTVGSSLSSVLGPPVPEIKVKEEV 788
 P+ T P PIT TT P P+ T G+ + P V

Sbjct: 2847 TTTVTPTPTPTGTQT-PTTPTIT---TTTTVTPTPTPT--GTQTPTTPTITTTTTVTPTPT 2900

Query: 789 EPMDIMRPVSAVPLATNTVSPS 811
 P P + P T TV+P+

Sbjct: 2901 TPTGTQTPTTT-PITTTTTVTPT 2922

Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23
 Identities = 174/717 (24%), Positives = 243/717 (33%)

Query: 96 VVVRYPVQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
 V P P T + + TV T P TP + + P P PT P

Sbjct: 2321 VTPTPTPTGTQTPTTPTITTTTTVTPTPTPTGTQTPTTPTITTTTTVTPTPTPTGTQTPT 2380

Query: 155 A-PAPPSTLSLPPKVP-GQVTVMESSIPQASAI PVATISGQQGHPSNLHHIMTNVQMS 212
 P +T P P G T T + P T +G Q P+ TT V +

Sbjct: 2381 TTPITTTTTVTPTPTPTGTQTPTTPTITTTTTVTPTPTPTGTQT-PTTPTITTTTTVTPTPT 2439

Query: 213 IIRSNAPGP---PLHIGASHLERGAAAAA-VMSSSKVTTVLRTSQLPNAATAQPAVQHI 268
 + P P+ + P +++ +TT T T P I

Sbjct: 2440 PTPTGTQTPTTPTITTTTTVTPTPTPTGTQTPTTPTITTTTTVTPTPTPTGTQTPTTPTI 2499

Query: 269 IHQPIQSRPPVTTSSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328
 + P T P T + T +P T T T + T++ P

Sbjct: 2500 TTTTPTPTPTPTGTQTPTTPTITTTTTVTPTPTPTGTQTPTTPTITTTTTVT---PTPT 2556

Query: 329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
 Q P + TT P+ GT + T + T TP T PI

Sbjct: 2557 PTGTQTPTTPTITTTTTVTPTPTPTGTQTPTTPTITTTTTVTPTPTPTGTQTPTTPTITT 2616

Query: 386 TNTI-PSATTAGSVSHQTAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQIHTHTSP 443
 T T+ P T G+ + T P +T T+T P+ + T TT V P T T

Sbjct: 2617 TTTVTPTPTPTGTQTPTTPTITTTTTVTPTPTPTGTQTPTTPTITTTTTVTPTPTPTPTGT 2675

Query: 444 RIQPDYPAERSSLIPISGHRASPNPVMETRSDNRPSVPVQFYFL-PTYPPSAYPLAAH 502
 + P ++ + +P P +T + + P+ + PT P+

Sbjct: 2676 QTPTTPTITTTTTVT-----PTPTPTGTQTPT----TTPITTTTTVTPTPTPTG--TQTP 2724

Query: 503 TYTPITSSVS-TIRQYVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQG 560
 T TPIT++ + T P Q P + IT T V T Q T

Sbjct: 2725 TTTPTITTTTTVTPTPTPTGTQTPTTPTITTTTTVTPTPTPT--TGTQTPTTPTITTTTTVT 2782

Query: 561 IQPAPISTQGIQAPIGTPTGI---QPAPLGTQGIHSATPINTOGL---QPAPMGTOQQP- 613
 P P TQ PI T P P GTQ + TPI T P P GTQ P

Sbjct: 2783 PTPTPTGTQTPTTPTITTTTTVTPTPTPTGTQ-TPTTPTITTTTTVTPTPTPTGTQTPTT 2841

Query: 614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSIL 671
 P T+ V T P + P + T T T +Q+ +T ++ P+

Sbjct: 2842 TPITTTTTVTPTPTPTGTQTPTTPTITTTTTVTPTPTPTGTQTPTTPTITTTTTVTPTPT 2901

Query: 672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNNDQPTIAVP---PTAQPPPTIPTMI 728
 T P + TP +T + T P PT Q P T P

Sbjct: 2902 PTGTQTPTTPTITTTTTVTPTPTPTGTQTPTTPTITTTTTVTPTPTPTGTQTPTTPTITT 2961

Query: 729 AAASPPSQPAVALSTIPGAVPITPPITTTAAAPPSVTVGSSLSSVLGPPVPEIKVKEEV 788
 P+ T P PIT TT P P+ T G+ + P V

Sbjct: 2962 TTTVTPTPTPTGTQT-PTTPTIT---TTTTVTPTPTPT--GTQTPTTPTITTTTTVTPTPT 3015

Query: 789 EPMDIMRPVSAVPPLATNTVSPS 811
P P + P T TV+P+
Sbjct: 3016 TPTGTQTPTTT-PITTTTTVTPT 3037

Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23
Identities = 174/717 (24%), Positives = 243/717 (33%)

Query: 96 VVVRYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
V P P T + + T V T P TP + + P P PT P
Sbjct: 2390 VTPTPTPTGTQTPTTTTPIITTTTTVTPTPTPTGTQTPTTTPIITTTTTVTPTPTPTGTQTPT 2449

Query: 155 A-PAPPSTLSLPPKVP-GQVTVTMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212
P +T P P G T T + P T +G Q P+ TT V +
Sbjct: 2450 TTPITTTTTVTPTPTPTGTQTPTTTPIITTTTTVTPTPTPTGTQT-PTTTPITTTTTVTPT 2508

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTVLRLPTSQLPNAATAQPAVQHI 268
+ P P+ + P +++ +TT T T P I
Sbjct: 2509 PTPTGTQTPTTTPIITTTTTVTPTPTPTGTQTPTTTPIITTTTTVTPTPTPTGTQTPTTTPI 2568

Query: 269 IHQIQSRPPVTTSSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRLTSLIQHPFSA 328
+ P T P T + T +P T T T + T++ P
Sbjct: 2569 TTTTIVTPTPTPTGTQTPTTTPIITTTTTVTPTPTPTGTQTPTTTPIITTTTTVT---PTPT 2625

Query: 329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
Q P + TT P+ GT + T + T TP T PI
Sbjct: 2626 PTGTQTPTTTPIITTTTTVTPTPTPTGTQTPTTTPIITTTTTVTPTPTPTGTQTPTTTPIIT 2685

Query: 386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQIHTSP 443
T T+ P+ T G+ + T P +T T+T P+ + T TT V P T T
Sbjct: 2686 TTTVTPTPTPTGTQTPTTTPIITTTTTVTPTPTPTGTQTPTTTPIITTTTTVTPTPTPTGT 2744

Query: 444 RIQPDYPAERSSLIPISGHRASPNPVAMETRSNRPSPVPVQFYFL-PTYPPSAYPLAAH 502
+ P ++ + +P P +T + + P+ + PT P+
Sbjct: 2745 QTPTTTPIITTTTTVT-----PTPTPTGTQTPT---TTPITTTTTVTPTPTPTG--TQTP 2793

Query: 503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQG 560
T TPIT++ + T P Q P + IT TV T Q T
Sbjct: 2794 TTPITTTTTVTPTPTPTGTQTPTTTPIITTTTTVTPTPTPT--TGTQTPTTTPIITTTTTVT 2851

Query: 561 IQPAPISTQGIQAPIGTPIG---QPAPLGTQGIHSATPINTQGL---QPAPMGTOQQPQ- 613
P P TQ PI T P P GTQ + TPI T P P GTQ P
Sbjct: 2852 PTPTPTGTQTPTTTPIITTTTTVTPTPTPTGTQ-TPTTPIITTTTTVTPTPTPTGTQTPTT 2910

Query: 614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAAT-TVVQTHSQSASTNAPAQGSSPRPSIL 671
P T+ V T P + P + T T T +Q+ +T ++ P+
Sbjct: 2911 TPITTTTTVTPTPTPTGTQTPTTTPIITTTTTVTPTPTPTGTQTPTTTPIITTTTTVTPTPT 2970

Query: 672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNDQPTIAVP---PTAQPPPTIPTMI 728
T P + TP +T + T P PT Q P T P
Sbjct: 2971 PTGTQTPTTTPIITTTTTVTPTPTPTGTQTPTTTPIITTTTTVTPTPTPTGTQTPTTTPIIT 3030

Query: 729 AAASPPSQPAVALSTIPGAVPITPPITIAAAPPSVTVGSSLSSVLGPPVPEIKVKEEV 788
P+ T P PIT TT P P+ T G+ + P V
Sbjct: 3031 TTTVTPTPTPTGTQT-PTTTPIT---TTTTVTPTPTPT--GTQTPTTTPIITTTTTVTPTPT 3084

Query: 789 EPMDIMRPVSAVPPLATNTVSPS 811
P P + P T TV+P+
Sbjct: 3085 TPTGTQTPTTT-PITTTTTVTPT 3106

Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23
Identities = 174/717 (24%), Positives = 243/717 (33%)

Query: 96 VVVRYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
V P P T + + T V T P TP + + P P PT P
Sbjct: 2459 VTPTPTPTGTQTPTTTTPIITTTTTVTPTPTPTGTQTPTTTPIITTTTTVTPTPTPTGTQTPT 2518

Query: 155 A-PAPPSTLSLPPKVP-GQVTVTMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212
P +T P P G T T + P T +G Q P+ TT V +
Sbjct: 2519 TTPITTTTTVTPTPTPTGTQTPTTTPIITTTTTVTPTPTPTGTQT-PTTTPITTTTTVTPT 2577

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTVLRLPTSQLPNAATAQPAVQHI 268
+ P P+ + P +++ +TT T T P I
Sbjct: 2578 PTPTGTQTPTTTPIITTTTTVTPTPTPTGTQTPTTTPIITTTTTVTPTPTPTGTQTPTTTPI 2637

Query: 269 IHQIQSRPPVTTSSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRLTSLIQHPFSA 328
+ P T P T + T +P T T T + T++ P
Sbjct: 2638 TTTTIVTPTPTPTGTQTPTTTPIITTTTTVTPTPTPTGTQTPTTTPIITTTTTVT---PTPT 2694

Query: 329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
Q P + TT P+ GT + T + T TP T PI
Sbjct: 2695 PTGTQTPTTTPIITTTTTVTPTPTPTGTQTPTTTPIITTTTTVTPTPTPTGTQTPTTTPIIT 2754

769

P+ T P PIT TT P P+ T G+ + P V
 Sbjct: 3169 TTTVTPTPTGTGT-PTTTPIT---TTTTVTPTPTPT--GTQTPTTPTITTTTPTPT 3222

Query: 789 EPMDIMRPVSAVPPLATNTVSPS 811
 P P + P T TV+P+

Sbjct: 3223 TPTGTQTPTTT-PITTTTPTPT 3244

Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23
 Identities = 174/717 (24%), Positives = 243/717 (33%)

Query: 96 VVVRPYQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
 V P P T + + T V T P TP + + P P PT P

Sbjct: 3080 VTPTPTPTGTQTPTTPTITTTTPTPTPTGTQTPTTPTITTTTPTPTPTGTQTPT 3139

Query: 155 A-PAPPSTLSLPPKVP-GQVVTMESSIPQASAI PVATISGQQGHPSNLHHIMTNNVQMS 212
 P +T P P G T T + P T +G Q P+ TT V +

Sbjct: 3140 TTPITTTTPTPTPTGTQTPTTPTITTTTPTPTPTGTQT-PTTPTITTTTPTPT 3198

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTVLRPTSQLPNAATAQPAVQHI 268
 + P P+ + P +++ +TT T T P I

Sbjct: 3199 PTPTGTQTPTTPTITTTTPTPTPTGTQTPTTPTITTTTPTPTPTGTQTPTTPTI 3258

Query: 269 IHQPIQSRPPVTTSSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328
 + P T P T + T +P T T T + T++ P

Sbjct: 3259 TTTTPTPTPTGTQTPTTPTITTTTPTPTPTGTQTPTTPTITTTTPTPT---PTPT 3315

Query: 329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
 Q P + TT P+ GT + T + T TP T PI

Sbjct: 3316 PTGTQTPTTPTITTTTPTPTPTGTQTPTTPTITTTTPTPTPTGTQTPTTPTIT 3375

Query: 386 TNIT-PSATTAGSVSHQTAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQIHTSP 443
 T T+ P T G+ + T P +T T+ P + T TT V P T T

Sbjct: 3376 TTTVTPTPTGTQTPTTPTITTTTPTPTPTGTQTPTTPTITTTTPTPTPTPTGT 3434

Query: 444 RIQPDYPAERSSLIPISGHRASPNPAMETRSNRPVSVQFQYFL-PTYPPSAYPLAAH 502
 + P ++ + +P P +T + + P+ + PT P+

Sbjct: 3435 QTPTTPTITTTTPTPT---PTPTPTGTQTPT---TPTITTTTPTPTPTPTG--TQTP 3483

Query: 503 TYTPITSSVS-TIRQYVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQ 560
 T TPIT++ + T P Q P+ IT T V T Q T

Sbjct: 3484 TTTPTTTTPTPTPTGTQTPTTPTITTTTPTPTPTPTG--TGTQTPTTPTITTTTPT 3541

Query: 561 IQPAPISTQGIQAPIGTPTGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTQQPQ- 613
 P P TQ PI T P P GTQ + TPI T P P GTQ P

Sbjct: 3542 PTPTPTGTQTPTTPTITTTTPTPTPTGTQ-TPTTPTITTTTPTPTPTGTQTPTT 3600

Query: 614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQSSPRPSIL 671
 P T+ V T P + P + T T T +Q+ +T ++ P+

Sbjct: 3601 TPTTTTPTPTPTPTGTQTPTTPTITTTTPTPTPTGTQTPTTPTITTTTPTPTPT 3660

Query: 672 RKKPATDGAQKPKSEIHVSMATPVTVSMETVSNQNDQPTIAPV---PTAQPPPTIPTMI 728
 T P + TP +T + T P PT Q P T P

Sbjct: 3661 PTGTQTPTTPTITTTTPTPTPTGTQTPTTPTITTTTPTPTPTGTQTPTTPTIT 3720

Query: 729 AAASPPSQPAVALSTIPGAVPITPITTTIAAAPPSVTVGSSLSVLGPPVPEIKVKEEV 788
 P+ T P PIT TT P P+ T G+ + P V

Sbjct: 3721 TTTVTPTPTGTQT-PTTTPIT---TTTTVTPTPTPT--GTQTPTTPTITTTTPTPT 3774

Query: 789 EPMDIMRPVSAVPPLATNTVSPS 811
 P P + P T TV+P+

Sbjct: 3775 TPTGTQTPTTT-PITTTTPTPT 3796

Score = 313 (47.0 bits), Expect = 4.2e-23, P = 4.2e-23
 Identities = 169/695 (24%), Positives = 245/695 (35%)

Query: 96 VVVRPYQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
 V P P T + + T V T P TP + + P P PT P

Sbjct: 3655 VTPTPTPTGTQTPTTPTITTTTPTPTPTGTQTPTTPTITTTTPTPTPTGTQTPT 3714

Query: 155 A-PAPPSTLSLPPKVP-GQVVTMESSIPQASAI PVATISGQQGHPSNLHHIMTNNVQMS 212
 P +T P P G T T + P T +G Q P+ TT V +

Sbjct: 3715 TTPITTTTPTPTPTGTQTPTTPTITTTTPTPTPTGTQT-PTTPTITTTTPTPT 3773

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTVLRPTSQLPNAATAQPAVQHI 268
 + P P+ + P +++ +TT T T P I

Sbjct: 3774 PTPTGTQTPTTPTITTTTPTPTPTGTQTPTTPTITTTTPTPTPTGTQTPTTPTI 3833

Query: 269 IHQPIQSRPPVTTSSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328
 + P T P T + T +P T T T + T++ P

Sbjct: 3834 TTTTPTPTPTGTQTPTTPTITTTTPTPTPTGTQTPTTPTITTTTPTPT---PTPT 3890

Query: 329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
 Q P + TT P+ GT + T + T TP T PI
 Sbjct: 3891 PTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIIT 3950

Query: 386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQIHTHTSP 443
 T T+ P+ T G+ + T P +T T+T P+ + T TT V P T T
 Sbjct: 3951 TTTVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTPTGT 4009

Query: 444 RIQPDYPAERSSSLIPISGHRASPNFVAMETRSNRPSPVQFQYFL-PTYPPSAYPLAAH 502
 + P ++ + +P P +T + + P+ + PT P+
 Sbjct: 4010 QTPTTTPIITTTTIVT-----PTPTPTGTQTPT---TTPITTTTIVTPTPTPTG--TQTP 4058

Query: 503 TYTPITSSVS-TIRQYFVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQG 560
 T TPIT++ + T P Q P + IT T V T Q T
 Sbjct: 4059 TTTPIITTTTIVTPTPTPTGTQTPTTTPIITTTTIVTPTPTPT--TGTQTPTTTPIITTTTIVT 4116

Query: 561 IQPAPISTQGIQAPIGTPGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTQQPQP 614
 P P TQ PI T P P GTQ + TPI T P P GTQ P
 Sbjct: 4117 PTPTPTGTQTPTTTPIITTTTIVTPTPTPTGTQ-TPITTTPIITTTTIVTPTPTPTGTQTPT- 4174

Query: 615 EGKTSAVVLADGATIVANPISNPFSAAPAATTVVQTHSQSASTNAPAQGSSPRPSILRKK 674
 T+ + T+ P P T ++ ++N P + S+P+ S
 Sbjct: 4175 ---TTPITTT--TTVTPTPTPTGTQTGPPTHTSTAPIAELTTSNPPPESTPQTSRSTSS 4229

Query: 675 PATDGAKPKSEIH--VSMATPVTVSMETVSNQNDQOPTIAVPP-TAQQPP--PTIPTMIA 729
 P T+ S + + M + S T + T++ PP T PP PT T
 Sbjct: 4230 PLTESTTLLSTLPPAEMTSTAPPSTPTAPTTSGGHTLSPPPSTTTSPPGTPTRGTTTG 4289

Query: 730 AASPPSQPAVALSTI---PGAVPITPP--ITTIAAAP-PPSVTVGSSLSSVLGPPVPEI 782
 ++S P+ V +T P P++ P I T P P SV + L+ P E+
 Sbjct: 4290 SSSAPTPTSTVQTTTSAWTPPTPLSTPISIIRTTGLRYPSSVLICCVLNDTYYPAGEEV 4349

Score = 279 (41.9 bits), Expect = 1.8e-19, P = 1.8e-19
 Identities = 138/540 (25%), Positives = 194/540 (35%)

Query: 278 PVTTSNAIPPAVVATVSATRAQSPVITTTAAH-----ATDSALSRP--TLSIQHPPSAA 329
 P+TT+ + P T + T +P+ TTT T + + P T + P
 Sbjct: 1946 PITTITTTIVTPTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTP 2005

Query: 330 ISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILAT 386
 Q P + TT P+ GT + T + T TP T PI T
 Sbjct: 2006 TGTQTPTTTPIITTTTIVTPTPTPTGTQTPTTTPIITTTTIVTPTPTPTGTQTPTTTPIITTT 2065

Query: 387 NTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQIHTHTSPR 444
 T+ P+ T G+ + T P +T T+T P+ + T TT V P T T +
 Sbjct: 2066 TTVTPTPTPTGTQTPTTTPIITTTTIVTPTPTPTGTQTPTTTPIITTTTIVTPTPTPTGTQ 2124

Query: 445 IQPDYPAERSSSLIPISGHRASPNFVAMETRSNRPSPVQFQYFL-PTYPPSAYPLAAHT 503
 P ++ + +P P +T + + P+ + PT P+ T
 Sbjct: 2125 TPTTTPIITTTTIVT-----PTPTPTGTQTPT---TTPITTTTIVTPTPTPTG--TQTP 2173

Query: 504 YTPITSSVS-TIRQYFVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQGI 561
 TPIT++ + T P Q P + IT T V T Q T
 Sbjct: 2174 TTPITTTTIVTPTPTPTGTQTPTTTPIITTTTIVTPTPTPT--TGTQTPTTTPIITTTTIVTPT 2231

Query: 562 QPAPISTQGIQAPIGTPGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTQQPQP-- 613
 P P TQ PI T P P GTQ + TPI T P P GTQ P
 Sbjct: 2232 TPTPTGTQTPTTTPIITTTTIVTPTPTPTGTQ-TPITTTPIITTTTIVTPTPTPTGTQTPTTT 2290

Query: 614 PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSILR 672
 P T+ V T P + P + T T T +Q+ +T ++ P+
 Sbjct: 2291 PITTITTTIVTPTPTPTGTQTPTTTPIITTTTIVTPTPTPTGTQTPTTTPIITTTTIVTPTPTP 2350

Query: 673 KKPATDGAKPKSEIHVSMATPVTVSMETVSNQNDQOPTIAVP---PTAQPPPTIPTMIA 729
 T P + TP +T + T P PT Q P T P
 Sbjct: 2351 TGTQTPTTTPIITTTTIVTPTPTPTGTQTPTTTPIITTTTIVTPTPTPTGTQTPTTTPIITTT 2410

Query: 730 AASPPSQPAVALSTIPGAVPITPPITTIAAAPPPSVTVGSSLSSVLGPPVPEIKVKEEVE 789
 P+ T P PIT TT P P+ T G+ + P V
 Sbjct: 2411 TTVTPTPTPTGTQT-PTTTPIT---TTTTVTPTPTPT--GTQTPTTTPIITTTTIVTPTPT 2464

Query: 790 PMDIMRPVSAVPPLATNTVSPS 811
 P P + P T TV+P+
 Sbjct: 2465 PTGTQTPTTT-PITTTTIVTPT 2485

Score = 265 (39.8 bits), Expect = 5.8e-18, P = 5.8e-18
 Identities = 179/746 (23%), Positives = 257/746 (34%)

Query: 96 VVVRYPVQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
 V P P T + + T V T P TP + + P P PT P
 Sbjct: 3678 VTPTPTPTGTQTPTTTPIITTTTIVTPTPTPTGTQTPTTTPIITTTTIVTPTPTPTGTQTPT 3737

Query: 155 A-PAPPSTLSLPPKVP-GQVVTMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212
P +T P P G T T + P T +G Q P+ T T V +
Sbjct: 3738 TTPITTTTTPPTPTGTQTPTTTPITTTTTPPTPTGTQT-PTTTPITTTTTPPT 3796

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTVLRLPTSQLPNAATAQPAVQHI 268
+ P P+ + P +++ +TT T T P I
Sbjct: 3797 PTPTGTQTPTTTPITTTTTPPTPTGTQTPTTTPITTTTTPPTPTGTQTPTTTP 3856

Query: 269 IHQPIQSRPPVTTSSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328
+ P T P T + T +P T T T + T++ P
Sbjct: 3857 TTTTTPPTPTPTGTQTPTTTPITTTTTPPTPTGTQTPTTTPITTTTTPPTPT 3913

Query: 329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
Q P + TT P+ GT + T + T TP T PI
Sbjct: 3914 PTGTQTPTTTPITTTTTPPTPTGTQTPTTTPITTTTTPPTPTGTQTPTTTP 3973

Query: 386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQIHTHTSP 443
T T+ P+ T G+ + T P +T T+T P+ + T TT V P T T
Sbjct: 3974 TTTTTPPTPTPTGTQTPTTTPITTTTTPPTPTGTQTPTTTPITTTTTPPTPT 4032

Query: 444 RIQPDYPAERSSLIPISGHRASPNPVAMETRSNRPSPVPVQFYFL-PTYPPSAYPLAAH 502
+ P ++ + +P P +T + + P+ + PT P+
Sbjct: 4033 QTPTTTPITTTTTPPTPTPTGTQTPTTTPITTTTTPPTPTPTGT--TQTP 4081

Query: 503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQG 560
T TPIT++ + T P Q P + IT T V T Q T
Sbjct: 4082 TTTTPTTTPPTPTPTGTQTPTTTPITTTTTPPTPTPTPT--TGTQTPTTTPITTTTTP 4139

Query: 561 IQPAPISTQGIQAPIGTPTGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTTQQPQP 614
P P TQ PI T P P GTQ + TPI T P P GTQ P
Sbjct: 4140 PTPTPTGTQTPTTTPITTTTTPPTPTGTQT-PTTTPITTTTTPPTPTPTGTQTGPP 4198

Query: 615 EGKTSAVVLADGATIVANPISNPFSAAPA---ATTVVQTHSQSA-STNAPA---QGSSPRP 668
TS +A+ T +NP P S+ P +T+ T S + ST PA S+ P
Sbjct: 4199 T-HTSTAPIAELT--SNP--PPESSTPQTSRSTSSPLTESTLLSTLPPAIENTSTAPP 4253

Query: 669 SILRKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNNDOPTIAVPPTAQPPPTIPTMI 728
S T G S + +P + ++ PT + T T PT
Sbjct: 4254 STPTAPTSTSGGHTLSPPPSTTTSPPGTPTRGTTTSSSAPTSTVQTTTSAWT-PTPT 4312

Query: 729 AAASPPSQPAVALSTIPGAVPITPPITTTIAAAPPPSVTVGGSLSSVLGPPVPEIKVKEEV 788
++P L P +V I + AP V G+ + E
Sbjct: 4313 PLSTPSIIRTGLRPPSSVLICCVLNDTYYPAGEEV-YNGTYGDTCTYFVNCLSCTLEF 4371

Query: 789 EPMDIMRPVSAVPLATNTVSPSLALLANNLSMPTSDLPPGASPRKKPRKQOH 841
S P + +T +PS ++ S PT P P P +Q++
Sbjct: 4372 YNWSCPSTPSPTPTPSKSTPTPSKP--SSTPSKPTPGTKPPECDFDPPRQEN 4422

Score = 254 (38.1 bits), Expect = 8.7e-17, P = 8.7e-17
Identities = 167/697 (23%), Positives = 245/697 (35%)

Query: 115 SATPVAVTAPPAHLTPAVPLSFSEGLMKPPPK--PTMPSPR-PIAPAPPSTLSLPPKV-PG 170
S + T PP TP+ P + + PPP P+ P+ PI P P ST +LPP P
Sbjct: 1587 SPPTITTTTPPTTTTPSPPTTTT---TPPTTTTPSPPTTTPITP-PTSTTLTPPTTTPS 1642

Query: 171 QVTVTMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMSIIRSNAPGPPLHIGASHL 230
T + P + P T + + TT I + P PP +
Sbjct: 1643 PPPTTTTTPPTTTTPSPPTTTTPSPPTTTTTPPTTTTPSSPI--TTTTPSPTTMTTTPS 1700

Query: 231 PRGAAAAVMSSSKVTVLRLPTSQLPNAATAQPAVQHIHQPIQS-RPPVTTSSNAIPPAV 289
P SS +TT P+S + P P + PP TT +PP
Sbjct: 1701 P-----TTTPSSPITTTTTPSS---TTTPSPPTTMTTTPSPTTTPSPPTTMTTLPPT 1751

Query: 290 VATVSATRAQSPVITT-TAAHATDSALSRPTLSIQH----PPSAAISIQRPAQSRDVTTR 344
++ T P IT T + + + P + + + S + +P ++
Sbjct: 1752 TSSPLTTPLPPSITPTTFSPFSTTTTPPCVPLCNWTGWLDGKPNFHKEGGDTLIGD 1811

Query: 345 ITLPSHPALGTPKQQLHTMAQKTIFSTGTPVAAATVAPILATN-----TIPSATTAGS 397
+ P A + + ++ I G V ++ N IP A
Sbjct: 1812 VCGPGWAANISCRATMYP--DVPIGQLGQTVVCDVSVGLICKNEDQKPGGVIPMAFCLNY 1869

Query: 398 VSHQTAPTSTI--VTMTVPSSHSHATAVTTSNIPVAKVVPQQIHTHTSPRIQPDYPAERSS 455
+ Q TMT + + + T TT+ I V T T + P ++
Sbjct: 1870 EINVCCECVTQPTTMTT---TENPTPTTTPITTTTTPPTPTPTGTQTPTTTPITTTT 1928

Query: 456 LIPISGHRASPNPVAMETRSNRPSPVPVQFYFL-PTYPPSAYPLAAHTYTPITSSVS-T 513
+ P P +T + + P+ + PT P+ T TPIT++ + T
Sbjct: 1929 TPTPTGTQTPTPT---TTPITTTTTPPTPTPTGT--TQTPTTTPITTTTTP 1977

Query: 514 IRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQGIQAPISTQGIQ 572

P Q P + IT T V T Q T P P TQ
 Sbjct: 1978 PTPPTGTQTPTTTTPIITTTTPTPTPT--TGTQTPTTTTPIITTTTPTPTPTGTQTPT 2035
 Query: 573 PAPIGTGPI---QPAPLGTQGIHSATPINTQGL---QPAPMGTTQQPQ---PEGKTSAVVLA 624
 P I T P P GTQ + TPI T P P GTQ P P T+ V
 Sbjct: 2036 TTPITTTTPTPTPTGTGTQ-TPTTPTITTTTPTPTPTGTGTPTTPTITTTTPTPT 2094
 Query: 625 DGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSILRKKPATDGAKEP 683
 T P + P + T T T +Q+ +T ++ P+ T P
 Sbjct: 2095 PTPGTQTPTTPTITTTTPTPTPTGTGTPTTPTITTTTPTPTPTGTGTPTTPTI 2154
 Query: 684 SEIHVSMATPVTVSMETVSNQNDQPTIAVP---PTAQPPPTIPTMIAAASPPSQPAVA 740
 + TP +T + T P PT Q P T P P+
 Sbjct: 2155 TTTTPTPTPTPTGTGTPTTPTITTTTPTPTPTGTGTPTTPTITTTTPTPTPTGT 2214
 Query: 741 LSTIPGAVPIPTPITTTIAAAPPSPVTVGSSLSSVLGPPVPEIKVKEVEPMDIMRPVSAV 800
 T P PIT TT P P+ T G+ + P V P P+
 Sbjct: 2215 TQT-PTTPTIT---TTTPTPTPTPT--GTQTPTTPTITTTTPTPTPTGTGTPTT- 2267
 Query: 801 PPLATNTVSPS 811
 P T TV+P+
 Sbjct: 2268 PITTTTPTPT 2278

Score = 243 (36.5 bits), Expect = 1.3e-15, P = 1.3e-15
 Identities = 110/406 (27%), Positives = 154/406 (37%)

Query: 121 VTAP-PAHLTPAVPLSFSEGLMKPPKPTMPSRPIAPAPPSTLSLPPKVPGQVTVMESS 179
 +T P P TP+ P + + L P P+ P+ PP+T PP T + ++
 Sbjct: 1396 ITTSPPTTTTSPPTTTTTL-PPTTTPSPPTTTTTPPTTTTSPPIIT--TTTLPPT 1452
 Query: 180 IPQASAI PVATISGQQGHPSNLHHIMTTNVQMSIIRSNA GPPLHIGASHLPRGAAAAV 239
 P P++T + P+ TT + P PP + P
 Sbjct: 1453 TPSP---PISTTTTPP--PTTTPSPPTTTSPSP--TTTSPPTTTTTPP-----TT 1498
 Query: 240 MSSSKVTTLVLRP---TSQLPNAATAQPAVQHIIHQPIQSRP-PVTTNAIPPAVVATVSA 295
 S +TT + P T+ LP T P P + P P TT+ PP T+
 Sbjct: 1499 TPSPPTTPTTTPASTTTLPPTTTPSPPTTTTTPPTTTTSPPTTTTPTPTSTTTLP 1558
 Query: 296 TRAQSPVITTTAAHATDSALSRPTLSIQHPFSAAISQRPAPQSRDV-TTRITLPSHPALG 354
 T SP TTT + S PT + PP+ + P + TT T P P
 Sbjct: 1559 TTTSPPTTTTTPPTTTPSPPTTTTTPPTTTTTPPTTTTSPPTTTTTP--PPTT 1616
 Query: 355 TPKQQLHTMAQKTIFFSTGTPVAAATVAPILATNTIPSATTAGSVSHTQAPTSTIVMTVP 414
 TP T +T P T +P T TP TT S T P+ I T T P
 Sbjct: 1617 TPSPPTTPTITPTSTTTLP-PTTTPSPPTTTTTPPTTTTSPPTTTTTPSPPTTTTTP 1675
 Query: 415 SHSSHATA-VTTSNIPVAKVVPQQITHTSPIQPDYPAERSSLIPISGHRASPNPVMAMET 473
 ++ ++ +TT+ P + T SP P P ++ PS SP P M T
 Sbjct: 1676 PPTTTPSPPTTTTTPPTTTM---TTPSPPTTTPSPPTTTT-PSSTTTPSPPTTMTT 1730
 Query: 474 RSDNR-PSVPVQFYFLPTYPPSAYPLAHTYTPITSSVSTIRQYVSAQAPNS 526
 S PS P LP S+ PL T TP+ S++ PS P +
 Sbjct: 1731 PSPTTTPSPPTTTMTLPTTTSS-PL---TTTLPSPITPTTSPFSTTTTPTT 1780

Score = 189 (28.4 bits), Expect = 8.0e-09, P = 8.0e-09
 Identities = 92/374 (24%), Positives = 133/374 (35%)

Query: 439 THTSPRIQPDYPAERSSLIPISGHRASPNPVMAMETRSNDRPSVPVQFYF-LPTYPPSAY 497
 T + P P P ++ +P + + P PS P+ LPT PS
 Sbjct: 1398 TPSPPTTTPSPPTTTTTLPTTTPSPPTTTTTPPTTTTSPPTTTTTPPTTTTTPSP- 1456
 Query: 498 PLAHTYTPITSSVSTIRQYVSAQAPNSAITAQTVGVGASTVHLNFMQL-MTVDASHAR 556
 P++ T P T++ S P S T T +T PM +T AS
 Sbjct: 1457 PISTTTTTPPTTTPSPPTTTPSPPTTTPSPPTTTTTPPTTTPSPPTTTPITPPASTTT 1516
 Query: 557 HIQGIQAPISTQGIQAPIGTQGIHSATPINTQGLQAPMGTTQQPQPEG 616
 P+P +T P P TP +P T I P +T L P T P P
 Sbjct: 1517 LPPTTTPSPPTTTTTPPTTTP---SPPTTPI--TPPTSTTTLP---TTTSPSP 1566
 Query: 617 KTSAVVLADGATIVANPISNPFSAAPAATTVVQTHSQSASTNAP--AQGSSPRPSILRKK 674
 T+ T +P P + P+ T+ T +T +P ++P P+
 Sbjct: 1567 TTTT---PPPTTTPSP---PTTTPSPPTTTTTPPTTTPSPPTTTTTPPTTTPSP 1620
 Query: 675 PATDGAKEPKSEIHVSMATPVTVSMETVSNQNDQPTIAV-PPTAQPPPTIPTMIAA--A 731
 P T P + + P T + PT PPT P P I T
 Sbjct: 1621 PTTTPTPTPT--TTTLPPTTTPSPPTTTTTPPTTTPSPPTTTTTPSPPTTTTTPPT 1678
 Query: 732 SPSPQPAVALSTIPGAVPITPITTTIAAAPPSPVTVGSSLSSVLGPPV-----PEIKVK 785
 + PS P + P TP TT ++P + T S ++ PP P
 Sbjct: 1679 TTPSPPTTTPSPPTTTMTTTPSPPTTTPSPPTTTTTPSTTTTPSPPTTMTTTPSPPTTTPS 1738

Query: 786 EEVEPMDIMRPVSAVPPLATNTVSPSL 812

M + P + PL T + PS+
Sbjct: 1739 PFTTTMTLPPTTTSSPLTTPLPSI 1765

Score = 185 (27.8 bits), Expect = 1.6e-09, P = 1.6e-09
Identities = 71/270 (26%), Positives = 99/270 (36%)

Query: 563 PAPISTQGIQAPAPIGTPGIQAPAPLGTQGIHSATP---INTQGLQAPMGTQQPQ---PEG 616
P+P +T P P TP P T + + TP I+T P P T P P
Sbjct: 1422 PSPPTTTTTTPPTTTTPS-PPITTTTTPLPTTTTPSPPISTT-TTPPTTTTPSPPTTTTPSP 1479

Query: 617 KTSAVVLADGATIVANPISNPFSAAPAATTVVQTHSQSASTNAPAQSSPRPSILRKKPA 676
T+ T P + P +P TT + T S +T P SP + P
Sbjct: 1480 PTTTSPPTTTTTTPPTTTTP---SPPMTTPI-TPPASTTTLPPTTTTPSPPTTTTTTPPP 1535

Query: 677 TDGAKPKSEIHVSMATPVTVSMETVSNQNDQPTIAVPPTAQPPPTIPTMIAAASPPSQ 736
T P + TP+T T + P+ P T PPPT + PS
Sbjct: 1536 TTTTSPPT-----TTPITPPTSTTLPTTTTPS-PPPTTTTTTPPTTTTPSPPTTTTPSP 1588

Query: 737 PAVALSTIPGAVPITPFTTTIAAAPPSVTVGSSLSSVLGPPVPEIKVKEEVEPMDIMRP 796
P + +T P +PP TT PPP+ T ++ + PP + P P
Sbjct: 1589 PTITTTTPPTTTTPSPPTTT-TTTPPTTTTPSPPTTTPITPPTSTTLPTTTTPSP--PP 1645

Query: 797 VSAVPPLATNTVSPSLALLANLMSPTSDLPFGASP 832
+ P T T SP + T+ PP +P
Sbjct: 1646 TTTTTPPTTTTPSPPTTTTPSPPTTTTTPPTTTTP 1681

Score = 183 (27.5 bits), Expect = 3.4e-09, P = 3.4e-09
Identities = 91/390 (23%), Positives = 139/390 (35%)

Query: 326 PSAAISIQRPQSRDVTTR-ITLPSHPALGTPKQQLHTMAQKTIFSTGTPVAAATVAPIL 384
PS + P + T T PS P T T I +T TP+ T +P +
Sbjct: 1399 PSPPTTTSPPTTTTTLPTTTTPSPPTTTTTPPTTTTPSPPIITTTTLPPTTTTPSPI 1458

Query: 385 ATNTIPSATTAGSVSHQAPTSTIVTMTVPSHSSHATAVTTSNIP--VAKVVPQQITHTS 442
+T T P TT S T P+ T +P+ ++ TT+ P + P T T
Sbjct: 1459 STTTTPPTTTTPSPPT-TTTPSPPTTTTPSPPTTTTTPPTTTTPSPPTTTPITPPASTTTL 1517

Query: 443 PRIQPDYPAERSSLIPIGHRASP---NPVAMETRSDNR--SVVQFQYFLPTYPPSAY 497
P P ++ P SP P+ T + P + P T PP+
Sbjct: 1518 PPTTTTPSPPTTTTTPPTTTTPSPPTTTPITPPTSTTLPTTTTPSPPTTTTTPPTTTT 1577

Query: 498 PLAHTYTPITSSVSTIRQYPVSAQAPNSAITAQTVGVASTVHLNPMQL-MTVDASHAR 556
P T TP +++T P + +P T T +T P +T S
Sbjct: 1578 PSPPTTTTPSPPTITTTTPPTTTTPSP--TTTTPPTTTTPSPPTTTPITPPTSTTT 1634

Query: 557 HIQGIQAPAPISTQGIQAPAPIGTPGIQAPAPLGTQGIHSATPINTQGLQAPMGTQQPQPEG 616
P+P T P P TP P P T T T P P
Sbjct: 1635 LPPTTTTPSPPTTTTTPPTTTTPS--P-PTTTTPSPPTTTTTPPTTTTPSSPITTTTPSP 1691

Query: 617 KTSAVVLADGATIVANPISNPFSAAPAATTVVQTHSQSASTNAPAQSSPRPSILRKKPA 676
T+ + T ++PI+ + P+TT + +T +P SP + + P
Sbjct: 1692 PTTMTTPSPPTTTTPSPPTTT-TTTPSSTTTTPSPPTTMTTPSPPTTTTPSPPTTTMTTLPP 1749

Query: 677 TDGAKPKSEIHVSMATPVTVSMETVSNQNDQPTIAVPP 715
T + P + + P+++ T S + PT P
Sbjct: 1750 TTTSSPLT----TTLPPSITPPTFSPFSTTTPTTPCVP 1784

Score = 176 (26.4 bits), Expect = 1.8e-07, P = 1.8e-07
Identities = 101/402 (25%), Positives = 142/402 (35%)

Query: 345 ITLPSHPALGTPKQQLHTMAQKTIFSTGTPVAAATVAPILATNTIPSATTAGSVSHQAP 404
IT PS P TP T +T +P T P T P TT + T P
Sbjct: 1396 ITTTPSPPTT-TPSPPTTTTTLPTTTTPSPPTTTTTPPTTTTPSPPIITTTTLPPTTT 1454

Query: 405 TSTIVTMTVPSHSSHATAVTTNIPVAKVVPQQITHTSPIQPDYPAERSSLIPIGHR 463
+ I T T P ++ + TT+ + P P T +P P PI+
Sbjct: 1455 SPPISTTTTPPTTTTPSPPTTTTPSPPTTTTTP--PPTTTPSPPTTTPITPP- 1511

Query: 464 ASPNPVAMETRSDNRPSVVPVQFQYFLPTYPPSAYPLAHTYTPITSSVSTIRQYPVSAQA 523
AS + T PS P T PP+ P + T TPIT ST P + +
Sbjct: 1512 ASTTTLPTTT----PSPPTTTT---TTPPTTTT-SPPTTTPITPPTSTTLPTTTTPS 1563

Query: 524 PNSAITAQ---TGVGVAHVHLNPMQLMTVDASHARHIQGIQAPAPISTQGIQAPAPIGTP 579
P T T +T +P + T P+P +T P P TP
Sbjct: 1564 PPPTTTTTPPTTTTPSPPTTTTPSPPTITTTTPPTT-----TPSPPTTTTTPPTTTTP 1618

Query: 580 G-----IQAPAPLGTQGIHSAT---PINTQGLQAPMGTQQPQPEGKTSAVVLADGATIV 630
I P P T + T P T P P T P S +
Sbjct: 1619 SPPTTTPITP-PTSTTLPTTTTPSPPTTTTTPPTTTTPSPPTTTTTPSPPTTTTTPPP 1677

THIS PAGE BLANK (USPTO)

Query: 756 TIAAAPPPSVTVGGSLSSVLGPPVPEIKVKEEVEPMDIMRPVSAPVPLATNTVSPSLALL 815
T PPP+ T ++ PP + PP T P+ +
Sbjct: 1568 T-TTTPTPTTTTSPPTTTTSPPTITTTTTPPTTTTSPPTTTTTTTPPTTTTSPPTTTTPT 1626

Query: 816 ANNLSMPTSDLPFGASPRKKP 836
 S T+ LPP +P P
 Sbjct: 1627 TPPTS--TTTLPPTTTSPPPP 1645

Score = 131 (19.7 bits), Expect = 1.2e-03, P = 1.2e-03
 Identities = 112/492 (22%), Positives = 174/492 (35%)

Query: 96 VVVRYPYQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
 V P P T + + T V T P TP + + P P PT P
 Sbjct: 3977 VTPTPTPTGTQTPTTTTPTTTTPTPTPTPTGTQTPTTTTPTTTTPTPTPTGTQTPT 4036

Query: 155 A-PAPPSTLSLPPKVP-GQVVTMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212
 P +T P P G T T + P T +G Q P+ TT V +
 Sbjct: 4037 TTPITTTTPTPTPTPTGTQTPTTTTPTTTTPTPTPTGTQT-PTTTTPTTTTPTPTPT 4095

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTVLRPTSQLPNAATAQPAVQHI 268
 + P P+ + P +++ +TT T T P I
 Sbjct: 4096 PTPTGTQTPTTTTPTTTTPTPTPTGTQTPTTTTPTTTTPTPTPTGTQTPTTTPTI 4155

Query: 269 IHQPIQSRPPVTTSSNAIPPA--VVATVSATRAQSPVITTA--AHATDSALSRLTSIQH 324
 + P T P + T + T +P T T H + + + T S
 Sbjct: 4156 TTTTPTPTPTGTQTPTTTTPTTTTPTPTPTGTGTGPPHTSTAPIAELTTSNPP 4215

Query: 325 PPSAAISIQRPAS--RDVTTTRI-TLPSPALGTPKQQLHTMAQKTI FSTGTPVAAATVA 381
 P S+ R S + TT + TLP PA+ + T T + T T++
 Sbjct: 4216 PESSTPQTSRSTSSPLTESTTLLSTLP--PAI----EMTSTAPPSTPTAPTTSGGHTLS 4269

Query: 382 PILATNTIPSAT-TAGSVS-HTQAPTSTIVTMTVPSSHSHATAVTTSNIPVAKVVPQIT 439
 P +T T P T T G+ + + APT + V T S A T + P++ P I
 Sbjct: 4270 PPSSTTSPPGTPTRGTTGSSSAPTSTVQTTTS-----AWTPTPTPLS--TPSIIR 4321

Query: 440 HTSPRIQPDYPAERSSLIPISGHRASPNP-VAMETRSDN----RPSVPVQFYFLPTYP- 493
 T ++P YP+ ++ +P V T D S+ +++ + P
 Sbjct: 4322 TTG--LRP-YPSSVLICVLNDTYYPAGEEVYNGTYGDTCTFVNCSLCTLEFYNWSCPS 4378

Query: 494 -PSAYPLAHTYTPITSSVSTIRQYPVSAQAPNSAITAQTGVGVASTVHLNPMQLMTVDA 552
 PS P + + TP S S+ P P T L + T
 Sbjct: 4379 TFSPTTPSKS-TPTSPKPSSTPSKPTPGTKPPECDFDPPRQENETWWLCLDCFMATCKY 4437

Query: 553 SHARHIQGIQ----PAPISTQGIQAPIGTP 579
 ++ I ++ P P + G+QP + P
 Sbjct: 4438 NNTVEIVKVECEPPMPPTCSNGLQPVREDP 4468

Score = 117 (17.6 bits), Expect = 1.8e-02, P = 1.8e-02
 Identities = 41/156 (26%), Positives = 55/156 (35%)

Query: 710 TIAVPPTAQPPPTIPTMIAAASPPSQPAVALSTIPGAVPITPPITTIAAAPPSVTVGG 769
 T + P T PPPT T + + PS P +T P +PPITT P P+ T
 Sbjct: 1398 TSPPTTTPSPPTTTTTLPTTTTSPPTTTTTPPTTTPSPPIIT-TTTLPTTTTSP 1456

Query: 770 SLSSVLGPPVPEIKVKEEVEPMDIMRPVSAVPLATNTVSPSLALLANNLSMPTSDLPFG 829
 +S+ PP P P + P T T SP T+ PP
 Sbjct: 1457 PISTTTTTP-----PTTTPSPPTTTPSPPTTTPSPPTTTTTP-PPTTTPSPPM 1504

Query: 830 ASPRKKPRKQHVISTEEGDMMETNSTDDEKSTAKS 865
 +P P + T T+T +T S
 Sbjct: 1505 TTPITPPASTTTLPTTTTSPPTTTTTPPTTTTPS 1540

Score = 61 (9.2 bits), Expect = 1.6e-09, P = 1.6e-09
 Identities = 23/93 (24%), Positives = 41/93 (44%)

Query: 397 SVSHTQAPTSTIVTMTVPSSHSHATAVTTSNIPVAKVV----PQIHTHTSPRIQPDYPAE 452
 S++ + +T T+T+P+ + T TT+ P + V P+ S I D+P+
 Sbjct: 1257 SITTRPSTLTFTTTITLPTTPTSFTTTTTTTTTPTSSVLSTTPKLCLLWSDWINEDHPSS 1316

Query: 453 RSS---LIPISGHRASPNPVAMETRSDNRPSVPVQ 484
 S P G +P + E RS P + ++
 Sbjct: 1317 GSDDGDREPFDGVCAPEDI--ECRSVKDPHLSLE 1349

Score = 50 (7.5 bits), Expect = 8.0e-09, P = 8.0e-09
 Identities = 16/41 (39%), Positives = 19/41 (46%)

Query: 334 RPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKTI FSTGTP 374
 RP+ TT ITLP+ P T T T+ ST TP
 Sbjct: 1261 RPSLTLTFTT-ITLPTTPTSFTTTTTTTTTPTSSVLST-TP 1299

Score = 46 (6.9 bits), Expect = 5.4e-08, P = 5.4e-08
 Identities = 24/106 (22%), Positives = 37/106 (34%)

Query: 324 HPPSAAISIQRPASQSRDVTTRITLPSHPALGTPKQQLHTMAQKTI FSTGTPVAAATVAPI 383
 +PP A++ + +S T + P G Q A G I

Subjct: 1196 YPPGASVPTETCKSCVCTNSSQVVCREEGKILNQTDGAFICYWEICGPNGTVEKHFNI 1255

Query: 384 LATNTIPSA-TTAGSVSHQAPTSTIVTMTVPSSHSHATAVTTNSI 428
+ T P S T T + + + P T S T T + + S T T +
Subjct: 1256 CSITTRPSTLTFTTTITLPTTPTSFTTTTTTTTTPTSTVLSTTPKL 1301

Score = 44 (6.6 bits), Expect = 8.7e-08, P = 8.7e-08
Identities = 14/34 (41%), Positives = 17/34 (50%)

Query: 478 RPSVPVQFQYF-LPTYPPSAYPLAAHTYTPITSSV 511
R P S F L P T P S + T T P + S + V
Subjct: 1261 RPSTLTFTTTITLPTTPTS-FTTTTTTTTTPTSTV 1294

Pedant information for DKFZphtes3_2a11, frame 2

Report for DKFZphtes3_2a11.2

[LENGTH] 1048
[MW] 110324.04
[pI] 9.83
[HOMOL] PIR:147141 gastric mucin (clone PGM-2A) - pig (fragment) 8e-15
[FUNCAT] 30.90 extracellular/secretion proteins [S. cerevisiae, YIR019c] 1e-09
[FUNCAT] 30.01 organization of cell wall [S. cerevisiae, YIR019c] 1e-09
[FUNCAT] 01.05.01 carbohydrate utilization [S. cerevisiae, YIR019c] 1e-09
[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YDR420w] 4e-09
[FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YDR420w] 4e-09
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YJR151c] 4e-06
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YGR014w] 1e-05
[FUNCAT] 11.01 stress response [S. cerevisiae, YHL028w] 1e-04
[FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YHL028w] 1e-04
[EC] 3.2.1.3 Glucan 1,4-alpha-glucosidase 3e-08
[PIRKW] glycosidase 3e-08
[PIRKW] transmembrane protein 3e-08
[PIRKW] polysaccharide degradation 3e-08
[PIRKW] glycoprotein 9e-08
[PIRKW] calcium binding 9e-08
[PIRKW] hydrolase 3e-08
[PIRKW] cytoskeleton 7e-08
[SUPFAM] equine herpesvirus glycoprotein X 2e-07
[SUPFAM] yeast glucan 1,4-alpha-glucosidase homolog 3e-08
[SUPFAM] polymorphic epithelial mucin 7e-08
[SUPFAM] glucan 1,4-alpha-glucosidase homology 3e-08
[SUPFAM] equine herpesvirus 1 glycoprotein homology 2e-07
[PROSITE] MYRISTYL 9
[PROSITE] AMIDATION 1
[PROSITE] CAMP_PHOSPHO_SITE 2
[PROSITE] CK2_PHOSPHO_SITE 10
[PROSITE] PKC_PHOSPHO_SITE 12
[PROSITE] ASN_GLYCOSYLATION 3
[KW] Irregular
[KW] LOW_COMPLEXITY 20.04 %

SEQ MGPPRHPQAGEIEAGGAGGRRRLQVEMSSQQFPRLGAPSTGLSQAPSQIANSAGSAGLINP
SEGXXXXXXXXXXXXX.....
PRD CCC

SEQ AATVNDESGRDSEVSAREHMSSSSLQSREEKQEPVVVRYPYQVQMLSTHHAVASATPVA
SEGXXXXX.....XXXXXXXXXXXXX
PRD CCC

SEQ VTAPPAHLTPAVPLSFSEGLMKPPPKPTMPSRPIAPAPPSTLSLPPKVPQVTVTMESSI
SEG XXXXXXXXXXXXXXXX.....XXXXXXXXXXXXX.....
PRD CCC

SEQ PQASAIIPVATISGQQGHPSNLHHIMTTNVQMSIIRSNAAGPPLHIGASHLPRGAAAAAVM
SEGXXXXX.....
PRD CCC

SEQ SSSKVTTVLRPTSQLPNAATAQPAVQHIHQPIQSRPPVTSNAIPPAVVATVSATRAQS
SEG
PRD CCC

SEQ FVITTTAAHATDSALSRPTLSIQHPFSAASIQRPAQSRDVTTRITLPSHPALGTPKQQL
SEG
PRD CCC

```

SEQ HTMAQKTI FSTGTPVAAATVAPILATNTIP SATTAGSVSHTQAPTSTIVTMTVPSHSSHA
SEG .....XXXXXXXXXX.....XXXXXXXX
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ TAVTTSNIPVAKVVPQQI THTSPRIQPDYPAERSSLIPI SGHRASPNPVMETRSDNRPS
SEG .....XXXXXXXX.....XXXXXXXX
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ VPVQFYFLPTYPPSAYPLAAHTYTPITSSVSTIRQYPVSAQAPNSAITAQTGVGVASTV
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ HLNPMQLMTVDASHARHIQGIQPAPISTQGIQPAPIGTPGIQPAPLGTQGIHSATPINTQ
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ GLQPAPMGTTQQPQPEGKTS AVVLADGATIVANPISNPFSAAPAATTVVQTHSQSASTNAP
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ AQGSSPRPSILRKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNDQPTI AVPPTAQQP
SEG .....XXXXXXXXXXXX
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ PPTIPTMIAAASPPSQPAVALSTIPGAVPITPPITTIAAAPPSVTVGSLSSVLGPPVP
SEG .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ EIKVKEEVEPMDIMRPVSAVPLATNTVSPSLALLANNLSMPTSDLP PGASPRKKPRKQ
SEG .....XXXXXXXXXXXX
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ HVISTEEDGMMETNSTDDEKSTAKSLLVKAERKSPKEYIDEEGVRYVPVRPRPPITLL
SEG .....XXXXXXXXXXXX
PRD cccccccccccccccccccccchhhhhhhhhcccccccccccccccccccccccccccccc

SEQ RHRYNPWKAAYHHFQRYSDVRVKEEKKAMLQEIANQKGVSCRAQGWKVLCAAQLQLTN
SEG .....
PRD eccccchhhhhhhccccchhhhhhhhhhhhhhhhhhhccccceeeccccceehhhhhhhhhc

SEQ LEHDVYERLTNLQEGII PKKKAATDDDLHRINELIQGNMQRCKLVM DQISEARDSMLKVL
SEG .....
PRD cchhhhhhhhhhhccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ DHKDRVLKLLNKNGTVKVSKLKRKEKV
SEG .....XXXXXXXXXXXX
PRD hhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccc

```

Prosites for DKFZphtes3_2a11.2

PS00001	818->822	ASN_GLYCOSYLATION	PDOC00001
PS00001	854->858	ASN_GLYCOSYLATION	PDOC00001
PS00001	1033->1037	ASN_GLYCOSYLATION	PDOC00001
PS00004	872->876	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	1037->1041	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	68->71	PKC_PHOSPHO_SITE	PDOC00005
PS00005	75->78	PKC_PHOSPHO_SITE	PDOC00005
PS00005	242->245	PKC_PHOSPHO_SITE	PDOC00005
PS00005	342->345	PKC_PHOSPHO_SITE	PDOC00005
PS00005	355->358	PKC_PHOSPHO_SITE	PDOC00005
PS00005	442->445	PKC_PHOSPHO_SITE	PDOC00005
PS00005	513->516	PKC_PHOSPHO_SITE	PDOC00005
PS00005	665->668	PKC_PHOSPHO_SITE	PDOC00005
PS00005	831->834	PKC_PHOSPHO_SITE	PDOC00005
PS00005	862->865	PKC_PHOSPHO_SITE	PDOC00005
PS00005	940->943	PKC_PHOSPHO_SITE	PDOC00005
PS00005	1035->1038	PKC_PHOSPHO_SITE	PDOC00005
PS00006	63->67	CK2_PHOSPHO_SITE	PDOC00006
PS00006	68->72	CK2_PHOSPHO_SITE	PDOC00006
PS00006	75->79	CK2_PHOSPHO_SITE	PDOC00006
PS00006	88->92	CK2_PHOSPHO_SITE	PDOC00006
PS00006	135->139	CK2_PHOSPHO_SITE	PDOC00006
PS00006	473->477	CK2_PHOSPHO_SITE	PDOC00006
PS00006	844->848	CK2_PHOSPHO_SITE	PDOC00006
PS00006	855->859	CK2_PHOSPHO_SITE	PDOC00006
PS00006	959->963	CK2_PHOSPHO_SITE	PDOC00006
PS00006	984->988	CK2_PHOSPHO_SITE	PDOC00006
PS00008	15->21	MYRISTYL	PDOC00008

PS00008	16->22	MYRISTYL	PDOC00008
PS00008	36->42	MYRISTYL	PDOC00008
PS00008	233->239	MYRISTYL	PDOC00008
PS00008	372->378	MYRISTYL	PDOC00008
PS00008	533->539	MYRISTYL	PDOC00008
PS00008	535->541	MYRISTYL	PDOC00008
PS00008	590->596	MYRISTYL	PDOC00008
PS00008	768->774	MYRISTYL	PDOC00008
PS00009	19->23	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3_2a11.2)

DKFZphtes3_2a17

group: metabolism

DKFZphtes3_2a17 encodes a novel 574 amino acid protein without similarity to known proteins.

The novel protein contains a thiol protease cys pattern. Eukaryotic thiol proteases (EC 3.4.22.-) are a family of proteolytic enzymes containing an active site cysteine. Cathepsins belong to this protease family.

The new protein can find application in modulation of proteolytic processes and as a new enzyme for proteomic analysis and biotechnologic production processes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 2312 bp

Poly A stretch at pos. 2300, polyadenylation signal at pos. 2273

```
1 GTTTTCACCT GATCATTAGA AACTAATGAA ACACCTTTTA AGTCTTATGA
51 ATTACAGGTTA CACTGTTTTC CAGATGCCTT GGCAGCTGGT ACAGGGCCTC
101 TGAAAAATGG AACCAAATTC TCTGAGGACT AAAGTCCCAG CTTTCTTATC
151 TGATTTGGGG AAGGCCACAT TGAGGGGAAT CAGAAAGTGT CCCCAGATGTG
201 GCACATACAA TGGAACCCGG GGAAGTGGCT GTAAGAACA GACATGTGGA
251 ACCATATTCC GCTACGGTGC ACGCAAGCAG CCTAGTGTG AAGCTGTCAA
301 AATCATTACA GGCTCTGATC TTCAGTCTA CTCAGTGCGG CAAAGAGACC
351 GGGGCCCTGA TTACCGATGC TTTGTGGAGC TCGGGGTTTC AGAGACAACA
401 ATCCAGACAG TGGATGGGAC GATCATCACT CAGCTGAGCT CTGGACGGTG
451 TTATGTCCCC TCATGCCTGA AAGCTGCCAC TCAAGGCGTT GTGGAAAAACC
501 AGTGCCAGCA CATCAAGCTG GCGGTGAACT GCCAGGCAGA GGCCACCCCT
551 CTGACCCCTGA AGAGCTCGGT CCTGAATGCA ATGCAGGCCT CCCCAGAAAC
601 CAAAACAGAC ATCTGGCAGT TGGCCACGGA ACCCAGAGT CCTCTGGTGC
651 AGAGAATTAC TAAAAACATC TTGGTGGTGA AATGCAAGGC AAGCCAGAAG
701 CACAGTTTGG GGTATTTGCA TACATCTTTT GTGCAGAAAG TCAGTGGCAA
751 AAGCTTGGCT GAGCGCCGCT TCTTCTGCTC CTGTCAGACT CTGAAATCGC
801 ACAAGTCAAA TGCCCTCAAAG GATGAGACAG CCCAGAGATG CATTCAATTC
851 TTTGCTTGA TCTGTGCTT TGCCAGTGAT GAGACACTGG CTCAGGAATT
901 CTCAGACTTC CTAAATTTT ATTCCAGCGG TCTTAAAGAG ATTATTGTAC
951 CCCAGTTAGG TTGCCATTCA GAATCAACAG TATCTGCTTG TGAGTCTACT
1001 GCCTCTAAGT CAAAGAAGAG GAGAAAGGAT GAAGTATCTG GTGCACAGAT
1051 GAACAGTTCA CTACTGCCTC AAGATGCAGT GAGCAGTAAT CTAAGGAAAA
1101 GTGGCCTGAA AAAGCCTGTG GTTGCTTCCT CGTTAAAAAG GCAGGCCTGT
1151 GGTACAGCTGT TAGATGAGGC ACAAGTGACT TTATCCTTCC AAGACTGGCT
1201 GGCCAGTGTC ACAGAACGCA TCCATCAAAC CATGCACTAT CAGTTTGATG
1251 GCAAACCCAGA ACCATTGGTG TTCCACATTC CTCAGTCATT TTTTGATGCC
1301 CTGCAACAAA GAATATCTAT AGGAAGTGCA AAAAAACGGC TCCCAACTC
1351 CACCACAGCT TTTGTTCGGA AAGATGCCTT GCCACTGGGA ACCTTTTCCA
1401 AGTATACTTG GCATATCACT AATATCCTGC AAGTTAAACA AATCTTAGAT
1451 ACCCCAGAGA TGCCCTTGGA AATCACCCTG AGCTTTATCC AGAACCAGAG
1501 TGGGACTTAT GAGCTATTTA AATGCCCTAA AGTGGAAAGT GAAAGCATAG
1551 CAGAAACCTA CGGTCGTATA GAAAAACAAC CAGTGCTGCG ACCCTTGAA
1601 CTAAAAACTT TTCTCAAAGT TGGCAACACT TCCCAGATC AAAAGGAGCC
1651 AACACCTTTC ATCATCGAGT GGATCCCAGA TATCCTTCCC CAATCTAAGA
1701 TTGGCGAGCT GCGGATCAAG TTTGAGTATG GCCACCACCG GAATGGGCAT
1751 GTGGCGGAGT ACCAAGACCA GCGGCCCCCC TTGGACCAGC CCTTGGAAGT
1801 GGCCCTCTG ACCACTATTA CTTTCCCTTA AAGCAAAACA AGATAATAAT
1851 CTTTGTCTGC TTAATTTGCA CATCCCACC CCTTGACAAC TTTAATGCT
1901 AGTTAGGCAC TTAGATGGCC CTGTTCTTGT GTAACTGCT CTTAGCTAAG
1951 ATGCAAAATC TCAGTGCTTT CAAGTGGATT CTGTTGAAGA AAATCTCTTG
2001 TAAATAGCCT TTTTGATGCT GCTGTGTACA GTCTTCATTA TGCATGGGC
2051 AGTATTTCTG GCTAGAGTTT TAAAAGGAAC AGAAAGAAAA CCAGCTTATT
2101 TTCCTTCTTA CGGACTCATC TTTAGCGTTT ATTTCAACCT TTGCTAATT
2151 CTCTGAGAAA TCTGCAGCAC TCAGCCATAC ACCAACAGTG TTGGAAGATT
2201 AACACCCCTG TTAGGGCAGA ATGTTAAAGA CCATCTTGGC AGAGTCCAG
2251 CCACGCTCTT TATTCTGTTC TCAATAAAG CAGTGTCACT AGTTTTCTCT
2301 AAAAAAAAAA AA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 107 bp to 1828 bp; peptide length: 574
 Category: putative protein

```

1  MEPNSLRITKV PAFSLDLGKA TLRGIRKCP RCGTYNGTRGL SCKNKTCTGI
51  FRYGARKQPS VEAVKIITGS DLQVYSVRQR DRGPDYRCFV ELGVSETTIQ
101 TVDGTIITQL SSGRCYVPSC LKAATQGVVE NQCQHIKLAV NCQAEATPLT
151 LKSSVLNAMQ ASPETKQTIW QLATEPTGPL VQRITKNILV VKCKASQKHS
201 LGYLHTSFVQ KVSQKSLPER RFFCSCQTLK SHKSNASKDE TAQRCIHFFA
251 CICAFADET LAQEFSDFLN FDSSGLKEII VPQLGCHSES TVSACESTAS
301 KSKRRRKDEV SGAQMNSSLL PQDAVSSNLR KSGLKPPVVA SSLKRQACGQ
351 LLDEAQVTL SFDWLASVTE RIHQTMHYQF DGKPEPLVFH IPQSFFDALQ
401 QRISIGSACK RLPNSTTAFV RKDALPLGTF SKYTWHITNI LQVKQILDTP
451 EMPLEITRSF IQNRDGTIEL FKCPKVEVES IAETYGRIEK QPVLRLPLEK
501 TFLKVGNTSP DQKEPTPFII EWIPDILPQS KIGELRIKFE YGHRNRGHVA
551 EYQDQRPPLD QPLELAPLTT ITFP

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2a17, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphtes3_2a17, frame 2

Report for DKFZphtes3_2a17.2

[LENGTH]	574
[MW]	64076.89
[pI]	9.15
[PROSITE]	MYRISTYL 5
[PROSITE]	CK2_PHOSPHO_SITE 9
[PROSITE]	PKC_PHOSPHO_SITE 14
[PROSITE]	ASN_GLYCOSYLATION 5
[PROSITE]	THIOL_PROTEASE_CYS 1
[KW]	Alpha_Beta


```

SEQ  MEPNSLRITKVPAFLSDLGKATLRGIRKCPRCGTYNGTRGLSCKNKTCTGIFRYGARKQPS
PRD  cccccccccchhhhhccccchhhhhcccccccccccccccccccccccccccccccccccccc

```

```

SEQ  VEAVKIITGSDDLQVYSVRQRDRGPDYRCFVELGVSETTIQTVDGTIITQLSSSGRCYVPSC
PRD  ceeeeeeccccccccccccccccccccccccccccccccccccccccccccccccccccccccch

```

```

SEQ  LKAATQGVVENQCQHIKLAVNCQAEATPLTLKSSVLNAMQASPETKQTIWQLATEPTGPL
PRD  hhhhhhhhhccccchhhhhheeehhhhhhhhccccchhhhhhhhhccccchhhhhhhhhccccch

```

```

SEQ  VQRITKNILVVKCKASQKHSGLYLHTSFVQKVSQKSLPERRFFCSCQTLKSHKSNASKDE
PRD  hhhhhhhheeecccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ  TAQRCIHFFACICAFADET LAQEFSDFLNFDSSGLKEIIVPQLGCHSESTVSACESTAS
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```

```

SEQ  KSKRRRKDEVSGAQMNSSLLPQDAVSSNLRKSGLKPPVVASSLKRQACGQLLDEAQVTL S
PRD  cccchhhhhccccccccccccccccccccchhhhhhhhhccccccccccccccccccccccccchhh

```

```

SEQ  FQDWLASVTERIHQTMHYQFDGKPEPLVFHIIPQSFFDALQQRISIGSACKRLPNSTTAFV
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```

```

SEQ  RKDALPLGTF SKYTWHITNI LQVKQILDTP E M P L E I T R S F I Q N R D G T Y E L F K C P K V E V E S
PRD  eccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccch

```

```

SEQ  IAETYGRIEKQPVLRLPLEK TFLKVGNTSPDQKEPTPFII EWIPDILPQSKIGELRIKFE
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```

SEQ YGHRNRNGHVAEYQDQRPPLDQPLELAPLTTITFP
 PRD eccccceeecccccccccccccccccccccccc

Prosites for DKFZphtes3_2a17.2

PS00001	35->39	ASN_GLYCOSYLATION	PDOC00001
PS00001	44->48	ASN_GLYCOSYLATION	PDOC00001
PS00001	235->239	ASN_GLYCOSYLATION	PDOC00001
PS00001	316->320	ASN_GLYCOSYLATION	PDOC00001
PS00001	414->418	ASN_GLYCOSYLATION	PDOC00001
PS00005	5->8	PKC_PHOSPHO_SITE	PDOC00005
PS00005	21->24	PKC_PHOSPHO_SITE	PDOC00005
PS00005	41->44	PKC_PHOSPHO_SITE	PDOC00005
PS00005	76->79	PKC_PHOSPHO_SITE	PDOC00005
PS00005	112->115	PKC_PHOSPHO_SITE	PDOC00005
PS00005	150->153	PKC_PHOSPHO_SITE	PDOC00005
PS00005	196->199	PKC_PHOSPHO_SITE	PDOC00005
PS00005	213->216	PKC_PHOSPHO_SITE	PDOC00005
PS00005	228->231	PKC_PHOSPHO_SITE	PDOC00005
PS00005	231->234	PKC_PHOSPHO_SITE	PDOC00005
PS00005	302->305	PKC_PHOSPHO_SITE	PDOC00005
PS00005	342->345	PKC_PHOSPHO_SITE	PDOC00005
PS00005	369->372	PKC_PHOSPHO_SITE	PDOC00005
PS00005	407->410	PKC_PHOSPHO_SITE	PDOC00005
PS00006	68->72	CK2_PHOSPHO_SITE	PDOC00006
PS00006	216->220	CK2_PHOSPHO_SITE	PDOC00006
PS00006	237->241	CK2_PHOSPHO_SITE	PDOC00006
PS00006	293->297	CK2_PHOSPHO_SITE	PDOC00006
PS00006	360->364	CK2_PHOSPHO_SITE	PDOC00006
PS00006	367->371	CK2_PHOSPHO_SITE	PDOC00006
PS00006	394->398	CK2_PHOSPHO_SITE	PDOC00006
PS00006	480->484	CK2_PHOSPHO_SITE	PDOC00006
PS00006	508->512	CK2_PHOSPHO_SITE	PDOC00006
PS00008	32->38	MYRISTYL	PDOC00008
PS00008	93->99	MYRISTYL	PDOC00008
PS00008	104->110	MYRISTYL	PDOC00008
PS00008	127->133	MYRISTYL	PDOC00008
PS00008	312->318	MYRISTYL	PDOC00008
PS00139	109->121	THIOL_PROTEASE_CYS	PDOC00126

(No Pfam data available for DKFZphtes3_2a17.2)

DKFZphtes3_2d15

group: testes derived

DKFZphtes3_2d15 encodes a novel 274 amino acid protein with similarity to C.elegans cosmid F25H2.1.

The novel protein contains a Pfam predicted C2-domain.
No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to C.elegans F25H2.1

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 3615 bp

Poly A stretch at pos. 3603, polyadenylation signal at pos. 3578

```
1 GCGGCGGCGCT CGAGGTGACA ACTGTCTCCG TCGCAGGCTC CGGCGGGGGC
51 GCAGGAGGTC GCCCGGCGCG TCACTGTCCG GTCGGCGAGC CACGGGGGCC
101 GCCGCAGCAC CATGGCGACC ACCGTGACGA CTCAGCGCGG GCCGGTGTAC
151 ATCGGTGAGC TCCCGCAGGA CTTCCTCCGC ATCAGGCCCA CACAGCAGCA
201 GCGGCAGGTC CAGCTGGACG CCCAGGCGGC CCAGCAGCTG CAGTACGGAG
251 GCGCAGTGGG CACCGTGGGC CGACTGAACA TCACGGTGGT ACAGGCAAAG
301 TTGGCCAAGA ATTACGGCAT GACCCGCATG GACCCCTACT GCCGACTGCG
351 CTTGGGCTAC GCGGTGTACG AGACGCCAC GGCACACAAT GCGCCAAGA
401 ATCCCGCTG GAATAAGGTC ATCCACTGCA CGGTGCCCC AGGCGTGGAC
451 TCTTTCTATC TCGAGATCTT CGATGAGAGA GCCTTCTCCA TGGACGACCG
501 CATTGCCTGG ACCCATCA CAATCCCGGA GTCCTGAGG CAGGGCAAGG
551 TGGAGGACAA GTGGTACAGC CTGAGCGGGA GGCAGGGGGA CGACAAGGAG
601 GGCATGATCA ACCTCGTCAT GTCCTACGCG CTGCTTCCAG CTGCCATGGT
651 GATGCCACCC CAGCCCGTGG TCCTGATGCC AACAGTGTAC CAGCAGGGCG
701 TTGGCTATGT GCCCATCACA GGGATGCCCG CTGTCTGTAG CCCCAGCATG
751 GTGCCCGTGG CCCTGCCCCC GCGCGCCGTG AACGCCACG CCCGCTGTAG
801 CGAGGAGGAC CTGAAAGCCA TCCAGGACAT GTTCCCCAAC ATGGACCAGG
851 AGGTGATCCG CTCCGTGCTG GAAGCCGAGC GAGGGAACAA GGATGCCGCC
901 ATCAACTCCC TGCTGCAGAT GGGGGAGGAG CCATAGAGCC TCTGCCTCGA
951 TGCCGTTTGG CCCCCTCTCT TTGGACACGC CGACCCGGCG CTCCCCAAGG
1001 AATGCTGTCC CAACAAGATT CCCGTGAAAG AGCACCCTGT TCGCCCCCTC
1051 CCGTGGACTT CTGTGCCGCC CCGTCCACAC CTGTTCTTGG GTGCATGTGG
1101 GTTTTCGGTT CCTGGCGGTC CAGGACGGGG CGGGGGCTCC CCTCCCATCT
1151 CGTGCTGGGA GGTCTCAGCG CGCTCTCTCT TCCCTGGGAC TGCCTCTCT
1201 CCTTCTCATG CCGTTCTGGA AAATGCTCTT GCTGTAGAGA GCAGTGCTT
1251 CTGCCAGGTT GTTGGAGGTG GTGGAGCGCC TTCCGATTCC ATTCATGGCA
1301 TTTTGTGATG TGATGTAATT GGAATAGAGC TGTGATTTA AGGCACACAC
1351 AATCCCTCAC ACTGTGGGTT TTTTGTAGAA CTCCCAGAC GAAACTCAC
1401 GCCCTTGCCC TAACGCGCTT TGCTGTGAGC CTGGCCCTG CCCAGGGCTT
1451 GGGTCTGGTG AGCTGAGCAG CTTCTGTGG ATGGTGTGG GCGGCCCTCT
1501 GGCCTGGCTC ACCTGGCCAC TGTCCAGCCA GCCTTGTGAC AGACTCCGGC
1551 CTGAAGGCAG AATGAACCCA CACCTGGAGT GAGGAAGGGG GCCTGGCAGG
1601 GTTGGCCAGG CTCTGCCTGA TTGCCAGCCA GCGGGCATCT GAAGCCGGGT
1651 CCTTCGCCCG CCGGAGGCTG CCGTCCGTCT CTCTGCTGC GCTCGTGCCA
1701 GCTCCGTGGG TGTCTCCCA GGGAGCTTCT CTCTCAACA GGCCTTGCGA
1751 GGCTGGGGTG AGAGGTGATA GAGGCAGCAC TGTGCATGAT TCCGAGAGGG
1801 TGTGTGGCA CTGCCAGCCG ACTGTGACA GCTTGGGAGC TGCTGTGCCC
1851 AGGACGTGGG TTCAGCGTGG GCGAGGAAAG CCTGGCGAGC GTGCCCTGT
1901 AAAAGCTTTC TGAGGCGGGA GCGCTCACT TACCTTGAC TGCCTGGGCG
1951 CTGCGTGTAG CATCTGGCC TACAGGACAG ATTTTAGGTG ACACCTGGTT
2001 ATGACAGTCA GAAATTGAG AAGCTTCTCA CAAGTGATGC ACTTTAATA
2051 ATCTGCATGC CATTGAGACA CCTGCATGTC TGGTGTGTGT GGTTCAGTG
2101 TCTTGCCGCC GGCCTTCGGA TGTAAACCCA CTGATAACGG ACAGAAGAG
2151 AATGCCACA AGTGGGTCTT CTGTGGAAGA TGCAGAAGGA GGAAGTTAGT
2201 GCTTACATTT TAGTCTTTT CTCCCTCAA AAAATAGGTT AAGTTTCAGT
2251 GCCAGCTAGA AAATACTGCT TTCTGCCATC GATTGGGGGT GGTTTTGTG
2301 AAATATACTG TTGATAAATA TTTATTTTGG TAAACTTGAA GTGTGTGGTG
2351 GCCGTGGGGG AGGGACATGC TGGCAGCAGG CGCCTTCTTC AGCTGTGGGT
2401 CCTAAGGCC TTTGATCCTT TGAAGAAGAA AGACATGGTA TTTGTTCAAG
2451 AGACGCCGAC CACTCAGACG GAGGGGCCCT TGGGATTCCC TGTCTCAGAT
2501 GGCCTGGTCT TACGCTGTG TAGATTCTT CTCCATTGGG AATGAAGGTG
2551 TCAGGCGGGA CTGGAACGTT CTAGATGGTA TGTCCGTGTA TATTAACAAC
2601 TCTAACCAG GACAGACCAC AAGCCACACT CAGAGGCCCT ACTGTGCTGG
```



```

2651 GGGCTTCGGT GTCCAGGCGC CCAGGTGTGG CCACCAGCAC CGGTTTCTGC
2701 CTTTCGCTTG CTGGGGTGCA GTGAGACTGC CACACGCTG CACATGTGGC
2751 TCTGTGGGTG TCTCCTAGAG AGGACGTGGC CCCTGCTGCC AGCCCTTGAG
2801 CAGCCCGTGT GGGGGCCCGA GGGACCCACA CAGTGGGGGC CAGCCTCGCT
2851 GGAGGGAGAG CAACCCCTTG CCGATGACCA CGCTTGCCGC CATCTCTTAG
2901 TTTTCTTTTT CACAAGCGCT TTATTTTTTT AATAGACAAA TCACATTTTG
2951 CAAGGCCTTT AATTAAATAA GATTCTTCTT TCCTTCATT TATGCTTTAT
3001 TTCCTGTTTG AAGGCTTACT GTAGAAGTGG CTTACTGTAG AAGCAGCTTG
3051 CTGAGCCCTT CCGAGCGGTC CCCAGAATTA GCTGGTTCAC AACCCCCACC
3101 CTCCCCCGCC CCCGCCGTG TCAGGTGTGG ATGAGGTCGT CACACTCAGA
3151 AGGACAGGCT TGTCTGCCAG CTCACAAGGG GAGGCTGCAG TGGGTTTGGG
3201 AGCTGGGTTT AGGCCCTTGG TGTCTGAGGG CCCAGGCCTT GCCAGCCTCT
3251 GCTGCTCCTG CTCCTGGGTT TGAAGATGCA GGCCGATCGC CAGCTCCGTG
3301 GCAGCGGTCA CTAAGGACAG CCTGACTGTG CCATCTTGA GCCTCAGCGC
3351 GGGCTCCGGA GATAGAAGAC AGGTCGCGGG AGGCTCCCCC TCCTCTCCTC
3401 TCCCCCTCTG AGATGCTCCC TGGGCGCTAC CCTGCAGGGT GCCAGGCAGG
3451 AGTGGTCTCA GAACGTGCGC TTCTGATTAT TTTACTGGGG TCCATTGTCC
3501 AGATTTTTCT TTGATTGTAA AATATATTTT TACTTTTTAG TCTTCTAATT
3551 TAATAAATGA TCCATATAAA AATAGAGAAA TAAAGTCCTT TAAGGGAAGG
3601 TTTAAAAAAA AAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 112 bp to 933 bp; peptide length: 274
 Category: similarity to unknown protein
 Classification: no clue

```

1 MATTVSTORG PVYIGELPQD FLRITPTQQQ RQVQLDAQAA QQLQYGGAVG
51 TVGRLNITVV QAKLAKNYGM TRMDPYCRLR LGYAVYETPT AHNGAKNPRW
101 NKVIHCTVPP GVDSFYLEIF DERAFSMDDR IAWTHITIFE SLRQGVDEK
151 WYSLSGRQGD DKEGMINLVM SYALLPAAMV MPPQPVVLMF TVYQQGVGYV
201 PITGMFAVCS PGMVPVALPP AAVNAQPRCS EEDLKAIQDM FPNMDQEVIR
251 SVLEAQRGNK DAAINSLQMQ GEEP

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFzphes3_2d15, frame 1

TREMBL:CEF25H2_1 gene: "F25H2.1"; Caenorhabditis elegans cosmid F25H2,
 N = 1, Score = 385, P = 1.1e-35

>TREMBL:CEF25H2_1 gene: "F25H2.1"; Caenorhabditis elegans cosmid F25H2
 Length = 457

HSPs:

Score = 385 (57.8 bits), Expect = 1.1e-35, P = 1.1e-35
 Identities = 77/182 (42%), Positives = 118/182 (64%)

```

Query:      4 TVSTQRGPVYIGELPQDFLRIT-PTQQQRQVQLDAQAAQQLQYGGAVGTVGRLNITVVQA 62
            TV+ +R V +GELP FLR+ P QQ + + + Q + + + T GRL++T+++A
Sbjct:      5 TVAERRRQVLVGELPPHFLRLAVPIQQTAPEPI-VQP-RMVSFVPP-NTRGRLSVTILEA 61

Query:      63 KLAKNYGMTRMDPYCRLRLGYAVYETPTAHNGAKNPRWNKVIHCTVPPGVDSFYLEIFDE 122
            L KNYG+ RMDPYCR+R+G ++T A N + P WN+ ++ +P V+S Y++IFDE
Sbjct:      62 NLVKNYGLVRMDPYCRVRVGNVEFDTNVAANAGRAPTWNRTL NAYLPMNVESIYIIFDE 121

Query:      123 RAFSMDDRIAWTHITIPESLRQGVDEKQWYSLSGRQDDKEGMINLVMSYAL--LPAAMV 180
            +AF D+ IAW HI +P ++ G D+++ LSG+QG+ KEGMI+L S+A LP
Sbjct:      122 KAFGPDEVIWAHIMLPLAIFNGDNIDEYFQLSGQQGEGKEGMIHLHFSFAPIDLPLQQA 181

```

Query: 181 MPPQP 185
P +P
Sbjct: 182 APAEP 186

Score = 92 (13.8 bits), Expect = 1.8e-01, P = 1.7e-01
Identities = 26/68 (38%), Positives = 38/68 (55%)

Query: 194 QQGVGYVPITGMPAVCSPGMVV--ALP--PAAVNAQPRCSEEDLKAIQDMFPNMDQEV 249
QQG G + + +P +P+ A P PA +EED K IQ+MFP +D+EVI
Sbjct: 156 QQGEGKEGMIHLHFSFAPIDLPLQQAAPAPAPLPVEITEEDTKEIQEMFPIVDKEVI 215

Query: 250 RSVLEAQR 257
+ +LE +R
Sbjct: 216 KCILEERR 223

Pedant information for DKFZphtes3_2d15, frame 1

Report for DKFZphtes3_2d15.1

[LENGTH] 274
[MW] 30281.97
[pI] 5.68
[HOMOL] TREMBL:CEF25H2_1 gene: "F25H2.1"; Caenorhabditis elegans cosmid F25H2 4e-36
[PFAM] C2 domain
[KW] Alpha_Beta
[KW] LOW_COMPLEXITY 16.42 %

SEQ MATTVSTQRGPVYIGELPQDFLRITPTQQQRQVQLDAQAAQQLQYGGAVGTVGRNLNITVV
SEGXXXXXXXXXXXXXXXXX.....
PRD cccccccccccccccccccccccccchhhhhhhhhhhhhhhcccccccccccccccc

SEQ QAKLAKNYGMTRMDPYCRLRLGYAVYETPTAHNGAKNPRWNKVIHCTVPPGVDSFYLEIF
SEG
PRD hhhhhhhhhccccccccchhhhhheeeeecccccccccccccccccccccccccccccccc

SEQ DERAFSMDDRIAWTHITIPESLRQGVKVEDKWYSLSGRQGDDEKGMINLVMSYALLPAAMV
SEGXXXXXXXXX.....
PRD ccc

SEQ MPPQPVVLMPTVYQQGVGYVPITGMPAVCSPGMVVPAALPAAVNAQPRCSEEDLKAIQDM
SEGXXXXXXXXX.....
PRD ccc

SEQ FPNMDQEVIRSVLEAQRGNKDAAINSLQMGEED
SEG
PRD cccccchhhhhhhhhhhccccchhhhhhhhhhhccc

(No Prosite data available for DKFZphtes3_2d15.1)

Pfam for DKFZphtes3_2d15.1

HMM_NAME C2 domain
HMM *LTVrIIeARNLWkMDMnGfSDPYVKVdMdPdpkDtkKWKtKiWNNGLN
L+++++A+ + + M+ DPY+++ + + + +T T +N N
Query 55 LNITVVQAKLAKNYGMT-RMDPYCRLRLGYAVY----ETPTAHNGAKN 97
HMM PVWNEEeFvFedIPyPdIqrkMLRFaVWDWDRFSRBDFIGHCi*
P+WN + +P + + ++++D+ FS +D I+ +
Query 98 PRWN-KVIHCT-VPPGVDSF---YLEIFDERAFSMDDDRIAWTH 135

DKFZphtes3_2e12

group: Transcription Factors

DKFZphtes3_2e12 encodes a novel 849 amino acid protein with similarity to Zinc finger proteins.

The new protein is a putative transcription factor with three C2H2 zinc fingers. Additionally, a cytochrome C family heme-binding site signature is present in the protein, which is only found in cytochrom C related proteins.

The new protein can find application in modulating/blocking the expression of genes controlled by this transcription factor.

similarity to finger proteins

complete cDNA, complete cds, 5 EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 3205 bp

Poly A stretch at pos. 3192, polyadenylation signal at pos. 3171

```

1 GGCACGGCCG GGTCTGGCT GGCCAAACGA GGCTCGCGGA AGCAGCAGCC
51 GCCGCTGAC CGCAGCTGGA TTTGAAGAT TGATCCAAGG GACTGTATTA
101 ATTTCAAGAA TTGATTGAA AGACACTGGC TCTGCCACTT AACAGCCATG
151 TAACCTTGA TATGGAAGAA AGTAGCAGTG TTGCCATGTT GGTGCCAGAT
201 ATTGGGGAAC AGGAAGCTAT ACTGACTGCT GAAAGTATCA TCAGTCTTTC
251 ATTGGAATTT GATGAACAAA GAAAACTAA ACCAGATCCA TTAATCCATG
301 TTATCCAGAA GTTAAGCAAG ATAGAAAAAT GAAAAGTCAC AAAAAATGCT
351 TTTAATTGGG AAGAAACGCC CACGTTCAAG TGCTGCAACA CACTCTCTTG
401 AAACCAAGA ACTTTGTGAG ATTCCGGCTA AAGTAATCCA GTCACCTGCT
451 GCTGATACTA GAAGGGCTGA GATGTCACAA ACAAATTTTA CCCTTGACAC
501 TCTTGCCAG AATGAAGGGA AGGCTATGTC TTATCAGTGT AGCCTTTGTA
551 AGTTTCTATC ATCATCCTTT TCCGTGTTAA AAGATCATAT TAAGCAACAT
601 GGTCAGCAAA ATGAAGTGAT ACTGATGTGC TCAGAGTGCC ATATTACATC
651 TAGAAGCCAG GAGGAACCTG AAGCCACGTT GGTGAATGAC CATGACAATG
701 ATGCCAATAT CCACACCCAA TCCAAAGCCC AACAGTGCGT AAGCCCTTCC
751 AGCTCTTTGT GTCGGAACAC CACAGAAAGA AATGAAACCA TTCCAGATAT
801 CCCAGTAAGT GTGGACAATC TACAGACTCA TACTGTCCAA ACTGCATCTG
851 TGGCAGAAAT GGGTAGGAGG AAATGGTATG CATACGAACA GTACGGCATG
901 TATCGATGCT TGTTTGTAG TTATACTTGT GGCCAGCAGA GAATGTTGAA
951 AACACACGCT TGGAAACATG CTGGGGAGGT TGATTGCTCC TATCCAATCT
1001 TTGAAAATGA AAATGAACCC CTAGGCCTGC TGGATTCTTC AGCAGCTGCT
1051 GCGCCTGGTG GGTTCGATGC AGTCGTCATT GCTATTGGAG AGAGTGAAC
1101 GAGTATCCAC AATGGGCCAT CAGTGCAAGT GCAGATTGTC AGCTCAGAAC
1151 AGTTATCATC TTCATCTCCT TTAGAACAGA GTGCAGAAAG AGGAGTACAC
1201 CTAAGTCACT CAGTTACCTT GGACCCCAAT GAGGAAGAAA TGCTAGAAGT
1251 GATTTCCTGT GCAGAGGAGA ATCTGATTCC TGATAGCCTG CTTACATCAG
1301 CACAGAAAAT CATCAGCAGC AGCCCCAATA AAAAAGGGCA TGTTAACGTG
1351 ATAGTGGAGC GATTGCCAAG TGCTGAAGAA ACCCTTTCAC AGAAGCGCTT
1401 CCTCATGAAC ACTGAAATGG AAGAAGGGAA GGACCTGAGC CTGACAGAAG
1451 CTCAGATTGG GCGCGAAGGA ATGGATGATG TTTATCGTGC TGATAAATGT
1501 ACTGTTGATA TTGGGGGATT GATCATAGGC TGGAGCAGTT CAGAGAAAAA
1551 AGACGAGTTA ATGAATAAAG GCCTGGCTAC TGATGAGAAAT GCCCCACCAG
1601 GCCGGAGAAG GACAAATTCT GAGTCTCTTC GATTACACTC ATTAGCTGCA
1651 GAAGCCCTTG TCACAAATGCC TATAAGAGCT GCAGAGTTGA CAAGAGCCAA
1701 CCTGGGGCAC TATGGAGATA TAAACCTTTT AGATCCAGAT ACTAGTCAAA
1751 GGCAAGTAGA TAGTACATTG GCAGCGTACT CAAAAATGAT GTCGCCACTT
1801 AAAAATCTTT CAGATGGATT AACTAGTCTT AACCAGGCA ACTCCACCTT
1851 GGTAGCACTC CCAGAGGGTA GGCAGGAATT GTCAGATGGG CAGGTTAAGA
1901 CAGGCATCAG CATGTCCTTA CTCACCGTCA TTGAAAAAAT GAGAGAAAGG
1951 ACAGACCAAA ACGCTTCAGA CGATGACATT TTGAAAGAGT TGCAGGACAA
2001 CGCCCAAGTG CAACCAACA GCGATACAAG TTTGTCCGGA AACAAATGTG
2051 TGGAATACAT CCCGAATGCT GAACGACCTT ACCGTTGCCG CCTGTGTGAC
2101 TACACAAGTG GCAACAAGGG CTACATCAAG CAGCACTTAC GAGTCCATCG
2151 ACAGAGACAG CCTTATCAGT GTCCTATCTG CGAGCACATA GCGGACACAA
2201 GCAAGATATT GGAGAGTCAC ATGATCCACC ACTGTAAGAC AAGAATATAC
2251 CAGTGAAGC AGTGTGAAGA ATCCTTCCAT TATAAGAGTC AATTGAGGAA
2301 CCATGAGAGA GAACAGCACA GTCTTCCAGA TACCTTGTCA ATAGCAACTT
2351 CTAATGAGCC AAGAATTTCC AGTGATACAG CTGATGGAAA ATGTGTCCAG
2401 GAAGGGAAAT AGTCTTCAGT CCAGAAACAA TATAGATGTG ATGTGTGTGA
2451 TTATACAAGT ACAACATATG TTGGTGTGAG AAACCACAGG CGAATCCATA
2501 ACTCTGATAA GCCGTACAGA TGCTCTCTGT GTGGGTATGT GTGTAGCCAT
2551 CCTCTTCTT TGAAGTCTCA TATGTGGAAA CATGCAAGTG ACCAAAAATTA

```

```

2601 CAACTACGAA CAAGTAAACA AGGCTATTAA CGACGCGATT TCACAAAGTG
2651 GCAGAGTTCT GGGGAAATCC CCTGGAAAGA CTCAATTAAA GAGCAGTGAA
2701 GAGAGTGCAG ATCCCGTCAC TGGAAAGTCG GAAAATGCAG TGTCATCTTC
2751 AGAACTGATG TCCCAGACTC CCAGTGAAGT TCTGGGTACC AACGAGAATG
2801 AGAAACTGAG CCTACAAGT AATACCTCAT ATAGTTTAGA AAAAATCTCC
2851 AGTCTGGCCC CTCCTAGCAT GGAGTACTGC GTTTTACTCT TCTGCTGTTG
2901 TATTTGTGTT TTTGAATCAA CCAGCAAAGA AAACCTCTTG GATCATATGA
2951 AAGAGCACGA GGGTGAAATT GTAAACATCA TCCTGAATAA GGACCACAAT
3001 ACAGCTCTAA ACACAAATTA GGTGGAATAA TGAATCGAGC AGGAAAGCAG
3051 TAGAAGAGGA TTCCTTCACC ACAGTTTCAC CTTTACGCTG TCAGACAACT
3101 TCCTGCCACA GAAGAAGTCG TTGATGTGAT TTTTGAGGAA ATGACAGATG
3151 TGACTTTGGA ACCAAACTTG TAATAAAAGG AATTCCAAAT GGAACAAAAA
3201 AAAAA

```

BLAST Results

No BLAST result

Medline entries

90301500:

Cloning and sequencing of a zinc finger cDNA expressed in mouse testis.

92310982:

Zfp-37, a new murine zinc finger encoding gene, is expressed in a developmentally regulated pattern in the male germ line.

Peptide information for frame 1

ORF from 472 bp to 3018 bp; peptide length: 849

Category: similarity to known protein

```

1 MSQTNFTPDT LAQNEGKAMS YQCSLCKFLS SSFSVLKDHI KQHGQQNEVI
51 LMCSECHITS RSQEELEAHV VNDHDNDANI HTQSKAQQCV SPSSSLCRKT
101 TERNETIPDI PVSVDNLQTH TVQTASVAEM GRRKWYAYEQ YGMYRCLFCS
151 YTCGQQRMLK THAWKHAGEV DCSYPIFENE NEPLGLLDSS AAAAPGGVDA
201 VVIAIGESL SIHNGPSVQV QICSSEQLSS SSPLEQSAER GVHLSQSVTL
251 DPNEEEMLEV ISDAEENLIP DSSLTSAQKI ISSSPNKKGH VNVIVERLPS
301 AEETLSQKRF LMNTEMEEGK DLSLTEAQIG REGMDDVYRA DKCTVDIGGL
351 IIGWSSSEKK DELMNKGLAT DENAPPGRRR TNSSELRHLS LAAEALVTMP
401 IRAAELTRAN LGHYGDINLL DPDTSQRQVD STLAAYSKMM SPLKNSSDGL
451 TSLNQSNSTL VALPEGRQEL SDGQVKTGIS MSLLTVIEKL RERTDQNASD
501 DDILKELQDN AQCQPNSDTS LSGNNVVEYI PNAERP YRCR LCHYTSGNKG
551 YIKQHLRVHR QRQPYQCPIC EHIADNSKDL ESHMIHCKT RIYQCKQCEE
601 SFHYKSQLRN HEREQHSLPD TLSIATSNEP RISSDTADGK CVQEGNKSSV
651 QKQYRCDVCD YTSTTYVGVR NHRRIHNSDK PYRCSLCGYV CSHPPSLKSH
701 MWKHASDQNY NYEQVNKAIN DAISQSGRVL GKSPGKTQLK SSEESADPVT
751 GSSENAVSSS ELSMQTPSEV LGTNEKLS PTSNTSYSLE KISSLAPPSM
801 EYCVLLFCCC ICGFESTSKE NLLDHMKHEH GEIVNIILNK DHNTALNTN

```

BLASTP hits

Entry S10245 from database PIR:

finger protein, testis - mouse

Score = 265, P = 8.4e-23, identities = 61/205, positives = 91/205

Entry S22954 from database PIR:

finger protein zfp-37 - mouse

Score = 265, P = 9.1e-22, identities = 61/205, positives = 91/205

Entry AF031657.1 from database TREMBL:

gene: "Zfp94"; product: "zinc-finger protein 94"; Rattus norvegicus zinc-finger protein 94 (Zfp94) gene, partial cds.

Score = 243, P = 1.6e-21, identities = 57/190, positives = 85/190

Alert BLASTP hits for DKFZphtes3_2e12, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_2el2, frame 1

Report for DKFZphtes3_2el2.1

```

[LENGTH]      849
[MW]           94325.42
[pI]           5.47
[HOMOL]        PIR:A54661 zinc finger protein ZNF41 - human (fragment) 2e-22
[FUNCAT]       04.05.01.04 transcriptional control [S. cerevisiae, YJL056c] 3e-09
[FUNCAT]       30.10 nuclear organization [S. cerevisiae, YJL056c] 3e-09
[FUNCAT]       04.03.01 trna synthesis [S. cerevisiae, YPR186c PZF1 - TFIIIA] 1e-07
[FUNCAT]       04.01.01 rna synthesis [S. cerevisiae, YPR186c PZF1 - TFIIIA] 1e-07
[FUNCAT]       04.99 other transcription activities [S. cerevisiae, YOR113w] 4e-07
[FUNCAT]       01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YGL209w]
2e-04
[FUNCAT]       13.04 homeostasis of other ions [S. cerevisiae, YNL027w] 2e-04
[FUNCAT]       11.01 stress response [S. cerevisiae, YMR037c] 3e-04
[BLOCKS]       BL00028 Zinc finger, C2H2 type, domain proteins
[SCOP]         dlmeyg_9.6.1.1.1 a designed zinc finger protein [syntheti 8e-06
[PIRKW]        nucleus 8e-18
[PIRKW]        RNA binding 5e-13
[PIRKW]        duplication 7e-13
[PIRKW]        tandem repeat 1e-21
[PIRKW]        spermatogenesis 6e-16
[PIRKW]        zinc 9e-21
[PIRKW]        zinc finger 1e-21
[PIRKW]        DNA binding 1e-21
[PIRKW]        metal binding 3e-15
[PIRKW]        phosphoprotein 5e-13
[PIRKW]        leucine zipper 1e-13
[PIRKW]        alternative splicing 6e-18
[PIRKW]        eye lens 2e-16
[PIRKW]        oocyte 1e-12
[PIRKW]        transcription factor 6e-18
[PIRKW]        segmentation 7e-13
[PIRKW]        embryo 1e-12
[PIRKW]        transcription regulation 2e-19
[PIRKW]        homeobox 2e-08
[SUPFAM]       POZ domain homology 7e-15
[SUPFAM]       transcription factor Krueppel 7e-13
[SUPFAM]       zinc finger protein ZFP-36 1e-21
[SUPFAM]       homeobox homology 2e-08
[SUPFAM]       unassigned homeobox proteins 2e-08
[PROSITE]      CYTOCHROME_C 1
[PROSITE]      MYRISTYL 10
[PROSITE]      ZINC_FINGER_C2H2 3
[PROSITE]      AMIDATION 2
[PROSITE]      CAMP_PHOSPHO_SITE 2
[PROSITE]      CK2_PHOSPHO_SITE 18
[PROSITE]      TYR_PHOSPHO_SITE 3
[PROSITE]      PKC_PHOSPHO_SITE 10
[PROSITE]      ASN_GLYCOSYLATION 7
[PFAM]         Zinc finger, C2H2 type
[KW]           Irregular
[KW]           3D
[KW]           LOW_COMPLEXITY 5.65 %

```

```

SEQ  MSQTNFTPTDLAQNEGKAMSQCSLCKFLSSSFVLKDHIKQHGQQNEVILMCSECHITS
SEG  .....XXXXXXXXXXXXXXXXX.....
lmeyF .....
SEQ  RSQEELEAHVVNDHDNDANIHTQSKAQCVSPSSSLCRKTTERNETIPDIPVSVNDNLQTH
SEG  .....
lmeyF .....
SEQ  TVQTASVAEMGRRKQYAYEQYGMRYCLFCSYTCGQQRMLKTHAWKHAGEVDCSYPIFENE
SEG  .....
lmeyF .....
SEQ  NEPLGLLDSSAAAAPGGVDVAVIAIGESLSIHNGPSVQVQICSSEQLSSSSPLEQSAER
SEG  .....XXXXXXXXXXXXXXXXX.....
lmeyF .....
SEQ  GVHLSQSVTLDPNEEEMLEVISDAEENLIPDSLLTSAQKIISSSPNKKGHVNVIVERLPS
SEG  .....
lmeyF .....

```

```

SEQ  AEETLSQKRFLMNTMEEGKDLSLTEAQIGREGMDDVYRADKCTVDIGGLIIGWSSEKK
SEG  .....
lme yF .....

SEQ  DELMNKGATDENAPPGRRTNSESRLHSLAAEALVTMPIRAAELTRANLGHYGDINLL
SEG  .....
lme yF .....

SEQ  DPDTSQRQVDSTLAAYSKMMSPLKNSSDGLTSLNQSNTLVALPEGRQELSDGQVKTGIS
SEG  .....
lme yF .....

SEQ  MSLLTVIEKLRERTDQNASDDDLKELQDNAQCQPNSDTSLSGNNVVEYIPNAERP YRCR
SEG  .....
lme yF .....TTTEETT

SEQ  LCHYTSGNKGYIKOHLRVHRQRPYQCPICEHIADNSKDLESHMIHCKTRIYQCKQCEE
SEG  .....
lme yF .....TTTCEETTHHHHHHHHHHTTCCEEETTTEEECHHHHHHHHHHHHCCCEETTTTE

SEQ  SFHYKSQLRNHEREQHSLPDTLSIATSNEPRISSDTADGKCVQEGNKSSVQKQYRCDVCD
SEG  .....
lme yF .....EECCHHHHHHHHHHHC.....

SEQ  YTSTTYGVNRHRRIHNSDKPYRCSLCGYVCSHPPSLKSHMWKHASDQNYNYEQVNKAIN
SEG  .....
lme yF .....

SEQ  DAISQSGRVLGKSPGKTQLKSSEESADPVTGSSSENAVSSSELMSQTPSEVLGTNENEKLS
SEG  .....
lme yF .....

SEQ  PTSNTSYLSLEKISSLAPPSMEYCVLLFCCICGFESTSKENLLDHMKHEGEIVNIILNK
SEG  .....
lme yF .....

SEQ  DHNTALNTN
SEG  .....
lme yF .....

```

Prosites for DKF2phtes3_2e12.1

PS00001	104->108	ASN_GLYCOSYLATION	PDOC00001
PS00001	445->449	ASN_GLYCOSYLATION	PDOC00001
PS00001	454->458	ASN_GLYCOSYLATION	PDOC00001
PS00001	457->461	ASN_GLYCOSYLATION	PDOC00001
PS00001	497->501	ASN_GLYCOSYLATION	PDOC00001
PS00001	646->650	ASN_GLYCOSYLATION	PDOC00001
PS00001	784->788	ASN_GLYCOSYLATION	PDOC00001
PS00004	98->102	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	378->382	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	59->62	PKC_PHOSPHO_SITE	PDOC00005
PS00005	101->104	PKC_PHOSPHO_SITE	PDOC00005
PS00005	306->309	PKC_PHOSPHO_SITE	PDOC00005
PS00005	357->360	PKC_PHOSPHO_SITE	PDOC00005
PS00005	385->388	PKC_PHOSPHO_SITE	PDOC00005
PS00005	425->428	PKC_PHOSPHO_SITE	PDOC00005
PS00005	678->681	PKC_PHOSPHO_SITE	PDOC00005
PS00005	696->699	PKC_PHOSPHO_SITE	PDOC00005
PS00005	726->729	PKC_PHOSPHO_SITE	PDOC00005
PS00005	817->820	PKC_PHOSPHO_SITE	PDOC00005
PS00006	62->66	CK2_PHOSPHO_SITE	PDOC00006
PS00006	106->110	CK2_PHOSPHO_SITE	PDOC00006
PS00006	126->130	CK2_PHOSPHO_SITE	PDOC00006
PS00006	232->236	CK2_PHOSPHO_SITE	PDOC00006
PS00006	262->266	CK2_PHOSPHO_SITE	PDOC00006
PS00006	300->304	CK2_PHOSPHO_SITE	PDOC00006
PS00006	314->318	CK2_PHOSPHO_SITE	PDOC00006
PS00006	323->327	CK2_PHOSPHO_SITE	PDOC00006
PS00006	355->359	CK2_PHOSPHO_SITE	PDOC00006
PS00006	381->385	CK2_PHOSPHO_SITE	PDOC00006
PS00006	485->489	CK2_PHOSPHO_SITE	PDOC00006
PS00006	499->503	CK2_PHOSPHO_SITE	PDOC00006
PS00006	617->621	CK2_PHOSPHO_SITE	PDOC00006
PS00006	626->630	CK2_PHOSPHO_SITE	PDOC00006
PS00006	741->745	CK2_PHOSPHO_SITE	PDOC00006
PS00006	758->762	CK2_PHOSPHO_SITE	PDOC00006
PS00006	766->770	CK2_PHOSPHO_SITE	PDOC00006
PS00006	817->821	CK2_PHOSPHO_SITE	PDOC00006

PS00007	331->339	TYR_PHOSPHO_SITE	PDOC00007
PS00007	703->711	TYR_PHOSPHO_SITE	PDOC00007
PS00007	596->605	TYR_PHOSPHO_SITE	PDOC00007
PS00008	142->148	MYRISTYL	PDOC00008
PS00008	185->191	MYRISTYL	PDOC00008
PS00008	196->202	MYRISTYL	PDOC00008
PS00008	241->247	MYRISTYL	PDOC00008
PS00008	349->355	MYRISTYL	PDOC00008
PS00008	473->479	MYRISTYL	PDOC00008
PS00008	478->484	MYRISTYL	PDOC00008
PS00008	645->651	MYRISTYL	PDOC00008
PS00008	751->757	MYRISTYL	PDOC00008
PS00008	772->778	MYRISTYL	PDOC00008
PS00009	130->134	AMIDATION	PDOC00009
PS00009	376->380	AMIDATION	PDOC00009
PS00028	146->167	ZINC_FINGER_C2H2	PDOC00028
PS00028	684->705	ZINC_FINGER_C2H2	PDOC00028
PS00028	595->617	ZINC_FINGER_C2H2	PDOC00028
PS00190	53->59	CYTOCHROME_C	PDOC00169

Pfam for DKFZphtes3_2e12.1

HMM_NAME Zinc finger, C2H2 type

HMM *CpwPDCgKtFrrwsNlRHMRT.H*

Query 53 CSE--CHITSRSQEELEAHVVN-DH 74

23.25 (bits) f: 539 t: 559 Target: dkfzphes3_2e12.1 similarity to finger proteins

Alignment to HMM consensus:

Query *CpwPDCgKtFrrwsNlRHMRT.H*

C C++T ++ ++H+R+H

dkfzphes3 539 CRL--CHYTSGNKGYIKQHLRVH 559

Query f: 567 t: 587 Target: dkfzphes3_2e12.1 similarity to finger proteins

Alignment to HMM consensus:

HMM *CpwPDCgKtFrrwsNlRHMRT.H*

CP+ C+ ++ +L+ HM+ H

Query 567 CPI--CEHIADNSKDLESHMIHH 587

33.47 (bits) f: 595 t: 616 Target: dkfzphes3_2e12.1 similarity to finger proteins

Alignment to HMM consensus:

Query *CpwPDCgKtFrrwsNlRHMRT.H*

C+ C+++F ++S+LR+H R H

dkfzphes3 595 CKQ--CEESFHYKSQLRNHERE-QH 616

Query f: 656 t: 676 Target: dkfzphes3_2e12.1 similarity to finger proteins

Alignment to HMM consensus:

HMM *CpwPDCgKtFrrwsNlRHMRT.H*

C++ C++T ++ R+H+R+H

Query 656 CDV--CDYTSTTYVGVNRHRIH' 676

24.53 (bits) f: 684 t: 704 Target: dkfzphes3_2e12.1 similarity to finger proteins

Alignment to HMM consensus:

Query *CpwPDCgKtFrrwsNlRHMRT.H*

C+ CG++ +++ +L+ HM H

dkfzphes3 684 CSL--CGYVCSHPPSLKSHMWKH 704

Query f: 809 t: 829 Target: dkfzphes3_2e12.1 similarity to finger proteins

Alignment to HMM consensus:

HMM *CpwPDCgKtFrrwsNlRHMRT.H*

C+ CG ++++NL HM+ H

Query 809 CCI--CGFESTSKENLLDHMKEH 829

DKFZphtes3_2f14

group: testes derived

DKFZphtes3_2f14 encodes a novel 129 amino acid protein with very weak similarity to human omega protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

weak similarity to omega protein

complete cDNA, complete cds, 1 EST hit

Sequenced by EMBL

Locus: unknown

Insert length: 2353 bp

Poly A stretch at pos. 2341, no polyadenylation signal found

```
1 GCAGATTCTC CAGGCCAGC ATCTGCCTCA CCGTGGCCCC CCACAAGCCA
51 AGCGCCTGCC TTTCAGCAGC CTCTACACAC CCAGCTCCTG CCACCCAATG
101 GCTCTTTAGG CCAAGCTCAT ACCTCACGAT GATTTTTCCTA GGCCCAACTT
151 TTGTCTCATG GCAACCTTCC CTGGCCAAGT TTCCACCTAT TTCCTGGCAG
201 CTGGGACAGG CCCAGGTCCT GCCACACACT GGCTCTCTTA CGCCAGCTC
251 ATGCCTCACA GTGGCCTCTC CAGGCCAGC TCCTGTCCCG GGACATCATC
301 TCCAGGCCCA AACTTCCTC AGTCGGCCT CTCCAGGCC AGTTGCTGCC
351 TCCCGGCATT CTCTCCAGGC CTAGCTCTTC CTCTGGCTG TATCTACAAG
401 ACCAACTCCT GCCTCACAAC AACCTTTTAT GGCTCAGCTC CTGCCCAACT
451 ACTGCCGGCC TTTGTAGGCC CAAAACCTCC TCAAGTCAAG CTCTTTAGGC
501 CCACCTTCTG CCTTGCAGTG GCCTGTACAG ACCAGCTCT GGCTTGAGAA
551 CAGCCTCTGC AGGCCCTGCT CTGCTCTCTT AGCTCCCTCT CCAGGCCCAT
601 CTCTTGCTC ACAGTGGCTT CCGTGGGCCA AGTTCGCCG TGCTCCCGAG
651 CAGCCTCAAC AGGCCTAGCT CCTCCCTCAC AATGGCTTGT TTAGGTCCAG
701 TTGATGCCCT TGGCAACCTG TCCAGGCCCA GCTCCTGCCT CACACTGGCC
751 TCTCTAGGCC GAGGTCTCTT CTCATACTGG CCTGTTTAGG CCCAGCTCAT
801 TCCTCTTGTC ATCTCTCCAG GCCCAGCTTT TGCTGTGTGT TGGCTCTAC
851 CTCACAGTGC ACCTTCCAGT CCCACCTCTT GCCTCACCAT GGCTCTCTCT
901 GACCAGGTTT CTGCTTTTCG GCAGCCTCTA CAGGCCTAGC TGCTGCCTCC
951 CAATGGCCTT TGTAGGCCAC GCTCATGCC CTCTGTGGCC TTTCAGGCC
1001 TAGCTTTCGC TTTTGGGCCA CTCCAGGCC AGAATTCCTC CCAGTCAGCC
1051 TCTCCAGGCC CAGCTCTTCC TCCAGCAAC CTCTGCAGGC CCAATCATC
1101 CTCAAATTGG CCTCTCTTCT CCCAGCTCCT GCCTCCTGGT GGCTCTGAA
1151 GAGCCAAATC GTCTCCAGT TGGTTTTTCC AGGCCAGCT CCTGCCTTTT
1201 GGTGGCCTCT CCAGGTGCAA AACTTCCTCC CATCAGCCTG TCCAGGCCCA
1251 GCTCATGCC CTGGGTGGCC TTCTCAGGCC CTGCTTTTGA CTTGGTGGCC
1301 TCTTCAGGCC CAGAACTTGA ACTCAAGTCA GCCTCTCCAG GCCCAGCTCC
1351 TGCCTTTCTA AGGTCTGTAC AGGCCAGCC TCTACCTCAC AGCGGACTCT
1401 CCACACCCAG CTCTTGCTC ACTGTAGCCT CCCAGTCCA AACTCCTGC
1451 CTTTTGGCAG CTTGACAAAG CCCAGCTCCT GCCTTTCAAT GACCTCTTTA
1501 GGCCCCGCTC ATTCTTACA ACGGCCTTTC CAGGCCAGT TTTTCCCTTT
1551 TGGCGGCTC TCCAGGCCCA GAATTCCTC AGTCGGCCT CTTTAGGCCC
1601 AGTTGCTGCC TCCTGGCATC CTCTGCAGGC CGAGCTCTTC CTCCTGCTG
1651 TGTCTACAGG CCCAACTCCT GCCTCACAAC AACCTCCTTG GACTCAGCTT
1701 CTGCCAGCT CCTGGTGGCC TTTGTAGGCT CAAAATTTTC TCAAATCAAG
1751 CTCTCCAGGC CTACTGTCAG CCTCGTGGCA GCCTAAACAG GCCCAGCTCC
1801 TGCCTGACAA TGGCCTCTCC AGGCTTTTCT CCTGCCTCGC AGCAGGCTTT
1851 CCAGGCCAG CTCTTGCTC ATGGTGGCCT TCCCGGCCA TGTTCCTATC
1901 TGACTTCTGG CAGCCTCAAC CGGCCAGCT TCTGCCTCAC ACTGGCCTCT
1951 CTAGGCCAG CTCTTTTTC ACAGTGGCCT CACTACGCC ATCTCCTACC
2001 TCAGATCTGC CTCCCAAGAC CCAGCTCCTG TCTCATGGTG GTCTCTCTTA
2051 CACCAGCTCC TGCCTCACA TGGCCTCGTC TGGCCCATCT TCTGCCTCAC
2101 AGTGGCCACT CAAGGCCCAT CTTTGTGCTC ATGGTAGCCT CTTCTGGTTT
2151 TGCTCTTGCC TCACAGTTGC CTCTTCCAGA TCCAGCTTTA AGCCTTTGAT
2201 GGTCAACAGC ATCAAGGAGC CTAAAGCTTC CTTGGACTCT CATTTGTTCA
2251 CTTTACAGCA GAGTGCCTTA GCAAAACTG TCTCTTAACC TTGAGAGTGG
2301 ATTTCTGACA AATCGATAGT AAATTCTGCC TGTGTGGTTT CAAAAAATAA
2351 AAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 158 bp to 544 bp; peptide length: 129
 Category: similarity to known protein

1 MATFPGQVST YFLAAWTGPG PATHWPLYAQ LMPHSGLSRP SSCPGTSSPG
 51 PKLPQVGLSR PSCCLPAFSP GLALPPGCIY KTNCLTTTF YGSAPAQLLP
 101 AFVGPKLPQV KLFRTFCLA VACTDPALA

BLASTP hits

Entry I70697 from database PIR:

omega protein - human (fragment)

Score = 79, P = 2.8e-03, identities = 32/94, positives = 38/94

Alert BLASTP hits for DKFZphtes3_2f14, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphtes3_2f14, frame 2

Report for DKFZphtes3_2f14.2

[LENGTH] 129
 [MW] 13421.76
 [pI] 9.14
 [PROSITE] MYRISTYL 2
 [KW] Irregular
 [KW] LOW_COMPLEXITY 10.85 %

SEQ MATFPGQVSTYFLAAWTGPGPATHWPLYAQLMPHSGLSRPSSCPGTSSPGPKLPQVGLSR
 SEGXXXXXXXXXXXXXXXXX.....
 PRD cccccccceehhhhhcc

SEQ PSCCLPAFSPGLALPPGCIYKTNCLTTTFYGSAPAQLLPAFVGPKLPQVKLFRPTFCLA
 SEG
 PRD ccccccccccccccccccccccccccecccccccccccccccccccccccccccccccccccc

SEQ VACTDPALA
 SEG
 PRD cccccccc

Prosite for DKFZphtes3_2f14.2

PS00008 6->12 MYRISTYL PDOC00008
 PS00008 92->98 MYRISTYL PDOC00008

(No Pfam data available for DKFZphtes3_2f14.2)

DKFZphtes3_2g7

group: testes derived

DKFZphtes3_2g7 encodes a novel 359 amino acid protein with similarity to neurofilament proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to neurofilament proteins

complete cDNA, complete cds, 6 EST hits (5 hits are out of a testis library)

Sequenced by EMBL

Locus: unknown

Insert length: 1613 bp

Poly A stretch at pos. 1595, polyadenylation signal at pos. 1557

```
1 GCCACACAGG CTCCTTGGAG TAAGAGTGTG AGAACTGGA TGAAGACAGC
51 TGTATTCTTT TGAAGCGTT CGAGATTGGT CTGTCTCTAC CAACTAAAAA
101 CTTCTAGCTT AAGTGCAGAG ATTTAAGGAG ATCAACAAAA ACTCAGTCTA
151 GACATATTAT GAGGCTGGGA GGGTATCAAC AGACTTGAGT TCTTGTGAGC
201 AAGATCACCT GCTTTTAATA TTGTCTCAG GGTCTGAGCA CATCTGGAAG
251 TGAGGTCAAT CAAGTTAGAC CCCAAAAACT TTTGTGACAA CAGTGAAGAG
301 GGGAAAAATA ACACACCACA AACATGAACC TCAACCCCCC GACATCTGCT
351 CTTCAGATCG AGGGCAAAGG CAGCCATATT ATGGCTAGAA ATGTAAGCTG
401 CTTTCTAGTC AGGCACACCC CTCATCCCAG AAGAGTCTGC CACATCAAAG
451 GCTTGAATAA CATTCCAATC TGTACTGTGA ATGATGATGA GAATGCATTT
501 GGAACATTGT GGAAGTTGG CCAGTCTAAC TACTTAGAGA AGAACAGGAT
551 ACCATTGGCC AATTGCAAGT ACCCCCCGAG CACTGCAGTC CAGAAGAGCC
601 CTGTAAGAGG AATGTCGCCA GCCCCAAACG GTGCCAAAGT GCCTCCACGG
651 CCTCATCTCG AGCCCAGTAG AAAAATTAAA GAGTGCTTCA AAACCTCCAG
701 TGAGAAATCCC TTAGTAATTA AAAAGGAAGA AATTAAGGCC AAAAGACCAC
751 CATCACCTCC AAAGGCATGC TCTACTCCTG GCTCCTGTTC TTCAGGGATG
801 ACAAGTACCA AGAATGATGT GAAAGCAAAC ACCATTGCA TACCAAACTA
851 TCTGGATCAG GAAATAAAAA TCCTGGCAAA GCTCTGTAGC ATTTTGATA
901 CTGATTCTCT GGCAGAGTT TTACAGTGGC TGCTTCATGC AACTTCAAAA
951 GAAAAAGAGT GGGTCTCAGC TTTGATTCAT TCTGAGCTTG CCGAGATAAA
1001 CCTGTAACT CATCACAGAA GAAACACCTC AATGGAACCA GCAGCAGAGA
1051 CTGGGAAGCC ACCCACAGTT AAATCACCAC CCACAGTTAA ATTGCCCCCA
1101 AATTTTACTG CAAATCAAAA AGTGCTGACC AGAGATACAG AAGGGGATCA
1151 ACCAACCAGA GTGTCAAGTC AAGGATCTGA AGAAAACAAG GAAGTACCAA
1201 AAGAGGCTGA GCACAAGCCT CCACTACTTA TAAGAAGAAA TAATATGAAA
1251 ATACCTGTTG CAGAATATTT CAGCAAACCA AATTCTCCTC CCAGGCCTAA
1301 CACTCAGGAG AGTGGATCAG CAAAACCAAG GTCAGCAAGG AGTATACAAG
1351 AATACAACCT CTGTCCCAA AGAGCATGTT ATCCTTCAAC ACACCGGAGG
1401 TAGAAGTTCT AGACTGGGTG AATTCTTTCA TGAATATGAG CTTACATTT
1451 ACATCATCAA ATTATTTTTC AAATGAATAT TTTTGGTATT GAGGAATCAA
1501 GTGGTCTCTT TTATGGTGGC ACATGTAAAT CTAAAATAC CTGTATGTAA
1551 TGCTACAAAT AAATATTACT GGAAATGATA TTTCCATTG TAGTTAAAAA
1601 AAAAAAAAAA AAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 324 bp to 1400 bp; peptide length: 359
Category: similarity to known protein

```

1 MNLNPPTSAL QIEGKGSHIM ARNVSCFLVR HTPHPRRVCH IKGLNNIPIC
51 TVNDDENAFG TLWEVGQSNY LEKNRIPFAN CSYPPSTAVQ KSPVRGMSPA
101 PNGAKVPPRP HSEPSRKIKE CFKTSSENPL VIKKEEIKAK RPPSPPKACS
151 TPGSCSSGMT STKNDVKANT ICIPNYLDQE IKILAKLCSI LHTDSLAEVL
201 QWLLHATSKE KEWVSALIHS ELAEINLLTH HRRNTSMEPA AETGKPPTVK
251 SPPTVKLPPN FTAKSKVLTR DTEGDQPTRV SSQGSEENKE VPKEAEHKPP
301 LLIRRNMMKI PVAEYFSKPN SPPRPNTQES GSAKPVARS IQEYNLCPQR
351 ACYPSTHRR

```

BLASTP hits

Entry A43427 from database PIR:

neurofilament triplet H1 protein - rabbit (fragment)

Score = 118, P = 5.6e-04, identities = 79/290, positives = 110/290

Entry RNNFH_1 from database TREMBL:

Rat high molecular weight neurofilament (NF-H) protein mRNA, 3' end.

Score = 115, P = 9.5e-04, identities = 69/281, positives = 100/281

Entry B43427 from database PIR:

neurofilament protein H form H2 (repetitive region) - rabbit (fragment)

Score = 111, P = 1.3e-03, identities = 64/269, positives = 102/269

Alert BLASTP hits for DKFZphtes3_2g7, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_2g7, frame 3

Report for DKFZphtes3_2g7.3

```

[LENGTH]      359
[MW]           39725.53
[pI]           9.45
[PROSITE]      MYRISTYL      3
[PROSITE]      CAMP_PHOSPHO_SITE      1
[PROSITE]      CK2_PHOSPHO_SITE      9
[PROSITE]      PKC_PHOSPHO_SITE     10
[PROSITE]      ASN_GLYCOSYLATION      4
[KW]           Alpha_Beta
[KW]           LOW_COMPLEXITY      4.18 %

SEQ  MNLNPPTSALQIEGKGSHIMARNVSCFLVRHTPHPRRVCHIKGLNNIPICTVNDDENAFG
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  TLWEVGQSNYLEKNRIPFANCYPPSTAVQKSPVRGMSPAPNGAKVPPRPHSEPSRKIKE
SEG  .....
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccchhhhhh

SEQ  CFKTSSENPLVIKKEEIKAKRPPSPPKACSTPGSCSSGMTSTKNDVKANTICIPNYLDQE
SEG  .....
PRD  hccccccccccccccccccccccccccccccccccccccccccccccccccccchhhh

SEQ  IKILAKLCSILHTDSLAEVLQWLLHATSKEKEWVSALIHSSELAEINLLTHHRRNTSMEPA
SEG  .....
PRD  hhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  AETGKPPTVKSPPTVKLPPNFTAKSKVLTRDTEGDQPTRVSSQGSEENKEVPKEAEHKPP
SEG  ....xxxxxxxxxxxxxxxx.....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  LLIRRNMMKI PVAEYFSKPN SPPRPNTQESGSAKPVARS IQEYNLCPQRACYPSTHRR
SEG  .....
PRD  eeeeecccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

Prosite for DKFZphtes3_2g7.3

```

PS00001      23->27  ASN_GLYCOSYLATION      PDOC00001
PS00001      80->84  ASN_GLYCOSYLATION      PDOC00001
PS00001     234->238  ASN_GLYCOSYLATION      PDOC00001

```

PS00001	260->264	ASN_GLYCOSYLATION	PDOC00001
PS00004	232->236	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	115->118	PKC_PHOSPHO_SITE	PDOC00005
PS00005	161->164	PKC_PHOSPHO_SITE	PDOC00005
PS00005	207->210	PKC_PHOSPHO_SITE	PDOC00005
PS00005	243->246	PKC_PHOSPHO_SITE	PDOC00005
PS00005	248->251	PKC_PHOSPHO_SITE	PDOC00005
PS00005	254->257	PKC_PHOSPHO_SITE	PDOC00005
PS00005	262->265	PKC_PHOSPHO_SITE	PDOC00005
PS00005	332->335	PKC_PHOSPHO_SITE	PDOC00005
PS00005	337->340	PKC_PHOSPHO_SITE	PDOC00005
PS00005	356->359	PKC_PHOSPHO_SITE	PDOC00005
PS00006	51->55	CK2_PHOSPHO_SITE	PDOC00006
PS00006	61->65	CK2_PHOSPHO_SITE	PDOC00006
PS00006	124->128	CK2_PHOSPHO_SITE	PDOC00006
PS00006	162->166	CK2_PHOSPHO_SITE	PDOC00006
PS00006	195->199	CK2_PHOSPHO_SITE	PDOC00006
PS00006	207->211	CK2_PHOSPHO_SITE	PDOC00006
PS00006	235->239	CK2_PHOSPHO_SITE	PDOC00006
PS00006	272->276	CK2_PHOSPHO_SITE	PDOC00006
PS00006	340->344	CK2_PHOSPHO_SITE	PDOC00006
PS00008	153->159	MYRISTYL	PDOC00008
PS00008	158->164	MYRISTYL	PDOC00008
PS00008	284->290	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_2g7.3)

DKFZphtes3_2h1

group: transmembrane protein

DKFZphtes3_2h1 encodes a novel 116 amino acid protein with weak similarity to C. elegans cosmid C13F10.

The novel protein contains 1 transmembrane region.
No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes and as a new marker for testicular cells.

similarity to C.elegans C13F10.5

TRANSMEMBRANE 1

Sequenced by EMBL

Locus: /map="2"

Insert length: 1156 bp

Poly A stretch at pos. 1143, polyadenylation signal at pos. 1121

```
1  GGCCATCAAA  ATAACATAAC  CATGTCATTT  GGAGCAACAA  AGCCACTGCG
51  GCCTCCATTT  GGGCCAAGCT  CTGACTGCAA  TGATGCCTCT  GCCCCGACCC
101  GGGCCTCGCT  GTGACTGACA  ATGCCGCTGC  ATCTTTTCAG  CAGTCATTGA
151  TGAGGAAGTA  TCTACATCCT  CTTCCCACT  ACCAGATTTT  GCTTGGAGAA
201  AAGCAGTTTC  CTGAAATAAT  TCTGTGACGA  GCTTCTTCCA  CATTAGGACA
251  AAAATGCTGG  AAGCGGCTCA  GCCCCAGGGC  AGCACATCAG  AGACACCATG
301  GAACACAGCC  ATTCTCTGTC  CGTCGTGCTG  GGACCAGTCT  TTCTGACCA
351  ATATCACCTT  CTTGAAGGTT  CTTCTCTGGT  TGGTCTGCT  GGGACTGTTT
401  GTGGAAGTGG  AATTTGGCCT  GGCATATTTT  GTCCTGTCTT  TGTCTATTG
451  GATGTACGTC  GGGACACGAG  GCCCTGAAGA  GAAGAAAGAG  GGAGAGAAGA
501  GCGCCTACTC  TGTGTTCAAT  CCAGGCTGTG  AAGCCATCCA  GGGCACCCTG
551  ACTGCAGAGC  AGTTGGAGCG  CGAGTTACAG  TTGAGACCCC  TGGCAGGGAG
601  ATAGGACCCA  GCTGTGCTGT  CATGCAGCTA  ACCTCTGATG  TGGTCTTCCT
651  CACCATTTGG  TATGGATTG  ATTTCAGGTG  TATAGGACTA  AGGGCAGCTT
701  GCGGGTTAGC  TCTGTGACTG  CATAGTTTTT  CTACCTTCTT  TCCCTGATCT
751  TTTGCTGCCA  TTTGATCTTT  GATAGTTTTG  GTGAAACTCT  CTAATAATACA
801  TTCACTGTGG  GTCCGACGCA  ATTTATAAAA  ATTATGTACT  CAAGAAGGGA
851  GACCTGTTTG  TTTCAATTCT  CATCTGTTTG  GGAGATGATT  TTAGAGCACT
901  AGAAAGGCAC  TGGGGAGATT  CTCAGCTTAA  AACATCCAGC  AGTTTGAAGT
951  ATGATTAGGT  ACATCAGGGC  TGCATTGTCA  ATGTTCTCTT  TAAGTCTTTT
1001  AACATTATTA  GCAATTTTTT  TTTTCCGGA  GAGTTTAGGT  TGCAAGTTTT
1051  GGGTTTCTTG  TTTGTTTTTG  TTTTGCTTCC  TGCTTTAATT  CTTTAATTTT
1101  CAGTCATTAC  TGGTATTGAA  AAATAAAATA  TCTTTAAAC  ATCAAAAAAA
1151  AAAAAA
```

BLAST Results

Entry HS313307 from database EMBL:
human STS SHGC-16715.
Score = 1222, P = 1.4e-48, identities = 248/251

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 254 bp to 601 bp; peptide length: 116
Category: similarity to unknown protein

```
1  MLEAAQPQGS  TSETPWNTAI  PLPSCWDQSF  LTNITFLKVL  LWLVLLGLFV
51  ELEFGLAYFV  LSLFYWMYVG  TRGPEEKKEG  EKSAYSVFNP  GCEAIQGTLT
101  AEQLERELQL  RPLAGR
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2h1, frame 2

TREMBL:CEUC13F10_2 gene: "C13F10.5"; Caenorhabditis elegans cosmid
C13F10., N = 1, Score = 141, P = 8.2e-10

>TREMBL:CEUC13F10_2 gene: "C13F10.5"; Caenorhabditis elegans cosmid
C13F10.
Length = 171

HSPs:

Score = 141 (21.2 bits), Expect = 8.2e-10, P = 8.2e-10
Identities = 32/82 (39%), Positives = 52/82 (63%)

Query: 27 DQSFLTNITFLKVLWLVLGLFVELEFGLAYFVLSLFYWMYVGTRGPPEKKEGEKSAYS 86
+QS ++ T + V++++V L ++FG +F+LSL + Y T G ++ GE SAYS
Sbjct: 90 EQSVVS--TRIAVVVVVVGQALAAWVQFQAVFFILSLILFTYWNT-G--RRRRGEMSAYS 144

Query: 87 VFNPGCEAIQGTLTAEQLEREL 108
VFN CE + G++TAE ER++
Sbjct: 145 VFNDNCERLAGSMTAEHFERDM 166

Pedant information for DKFZphtes3_2h1, frame 2

Report for DKFZphtes3_2h1.2

[LENGTH] 116
[MW] 13092.19
[pI] 4.64
[PROSITE] MYRISTYL 1
[PROSITE] CK2_PHOSPHO_SITE 2
[PROSITE] TYR_PHOSPHO_SITE 2
[PROSITE] ASN_GLYCOSYLATION 1
[KW] TRANSMEMBRANE 1
[KW] LOW_COMPLEXITY 32.76 %

SEQ MLEAAQPGSTSETPWNTAIPSCWDQSFLTNITFLKVLWLVLGLFVELEFGLAYFV
SEGXXXXXXXXXXXXXXXXXXXXX...
PRD cccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEMMMMMMMMMMMMMMMMMMM.....

SEQ LSLFYWMYVGTRGPPEKKEGEKSAYSVFNPGCEAIQGTLTAEQLERELQLRPLAGR
SEGXXXXXXXXXXXXXXXXXXXXX...
PRD hhhhhhhhhccccchhhhhccccccccccccccccccccchhhhhhhhhhhcccccc
MEM

Prosite for DKFZphtes3_2h1.2

PS00001	33->37	ASN_GLYCOSYLATION	PDOC00001
PS00006	10->14	CK2_PHOSPHO_SITE	PDOC00006
PS00006	24->28	CK2_PHOSPHO_SITE	PDOC00006
PS00007	78->86	TYR_PHOSPHO_SITE	PDOC00007
PS00007	77->86	TYR_PHOSPHO_SITE	PDOC00007
PS00008	97->103	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_2h1.2)

DKFZphtes3_2h15

group: testes derived

DKFZphtes3_2h15 encodes a novel 855 amino acid protein with very weak similarity to *S. pombe* cdc23.

No informative BLAST results; no predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to cdc23

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 4619 bp

Poly A stretch at pos. 4598, polyadenylation signal at pos. 4589

```
1 GAAGGCGTCC CGGCATCGGC CAAGATTCTA CATTGCTCAT CTGGGCATCT
51 GAGCCTCCTT CGAAGTTTCC TGTCACTACT GTCTCTTGA CAGCATGGAT
101 GAGGAGGAAG ACAATCTGTC TCTGCTGACC GCACTGCTGG AAGAAAATGA
151 GTCAGGCTTG GATTGTAATT CAGAAGAAAA TAACTTCTTG ACGCGGGAAA
201 ATGGCGAGCC CGACGCATT TATGAGCTCT TTGATGCCGA CGGCGACGGT
251 GAATCTTATA CAGAAGAGGC TGATGATGGA GAAACAGGAG AGACAAGAGA
301 CGAAAAGGAA AATCTGGCCA CTCTCTTTGG AGATATGGAG GACTTAACAG
351 ATGAAGAAGA AGTTCCCGCA TCACAGTCAA CTGAAAATAG GGTCTCCCT
401 GCTCTGCCCC CCAGGCGAGA GAAAACGAAT GAAGAGTTGC AAGAGGAATT
451 AAGGAATTTG CAAGAGCAAA TGAAGGCCTT ACAAGAGCAG CTAAAAGTAA
501 CAACAATTAA ACAGACAGCA AGCCCAGCCC GTCTGCAAAA ATCCCTGAG
551 AAGTCTCCCC GGCCACCTCT TAAGGAGAGG AGAGTTCAGA GAATTCAGGA
601 GTCAACATGC TTTTCTGCGG AGCTTGATGT CCCTGCGCTA CCAAGAACCA
651 AGAGGGTGGC TCGAACACCA AAGCCTTCAC CTCCAGATCC CAAAAGCTCA
701 TCTTCAAGGA TGACAAGTGC ACCCTCCCAA CCCCTACAGA CGATTTCTCG
751 GAACAAACCT AGTGGGATAA CTAGAGGTCA AATTGTGGGG ACCCCAGGAA
801 GTTCTGGGGA AACGACTCAA CCCATCTGTG TGGAAAGCCTT CTCTGGTCTG
851 CGGCTCAGGC GGCCTCGAGT ATCCTCCACA GAAATGAACA AGAAAATGAC
901 CGGCCGAAAA CTGATCAGAC TGTCTCAGAT CAAGGAAAAG ATGGCCAGAG
951 AGAAGCTGGA AGAAATAGAT TGGGTGACAT TTGGGGTTAT ATTGAAGAAG
1001 GTTACGCCAC AGAGTGTGAA TAGTGGAAAA ACCTTCAGCA TTGGGAAACT
1051 GAATGATCTT CGTGACCTGA CACAATGTGT GTCTTGTTC TATTTGGAG
1101 AAGTTCACAA AGCGCTCTGG AAGACGGAGC AGGGGACTGT CGTAGGGATC
1151 CTCATGCCA ACCCATGAA GCCCAAGGAT GGTTCAGAGG AGGTGTGTTT
1201 ATCTATCGAT CATCTCAGA AGGTCTTAAT TATGGGTGAA GCTCTTGACC
1251 TGGGAACCTG TAAAGCCAAG AAGAAGAATG GAGAGCCGTG CACGCAGACT
1301 GTGAATTTGC GTGACTGTGA GTACTGTGAG TACCATGTCC AGGCTCAGTA
1351 CAAGAAGCTC AGTGCAAAGC GTGCGGATCT GCAGTCCACC TTCTCTGGAG
1401 GACGAATTCC AAAGAAGTTT GCCCGCAGAG GCACCAGCCT CAAAGAACGG
1451 CTGTGCCAAG ATGGCTTTTA CTACGGAGGG GTTCTTCTG CCTCGTATGC
1501 AGCTTCAATT GCAGCAGCTG TGGCTCCTAA GAAGAAGATT CAAACCACTC
1551 TGAGTAATCT GGTGTGTAAG GGCACAAACT TGATCATCCA GGAACACCGG
1601 CAAAACCTCG GAATACCCCA GAAGAGCCTG TCTTGCTCTG AGGAGTTCAA
1651 GGAAGCTGAT GACCTGCCGA CGTGTGGAGC CAGGAACCTA AAACAACATT
1701 TAGCCAAAGC CTCAGCTTCA GGGATTATGG GGAGCCCAA ACCAGCCATC
1751 AAGTCCATCT CGGCCTCAGC ACTCTTGAAG CAACAGAAGC AGCGGATGTT
1801 GGAGATGAGG AGAAGGAAAT CAGAAGAAAT ACAGAAGCGA TTTCTGCAGA
1851 GCTCAAGTGA AGTTGAGAGC CCAGCTGTGC CATCTTCATC AAGACAGCCC
1901 CCTGCTCAGC CTCCACGGAC AGGATCCGAG TTCCCCAGGC TGGAGGGAGC
1951 CCCGGCCACA ATGACGCCCC AGCTGGGGCG AGGTGTCTTG GAAGGAGATG
2001 ATGTTCTCTT TTATGATGAG TCACCACCAC CAAGACCAAA ACTGAGTGCT
2051 TTAGCAGAAG CCAAAAAGTT AGCTGCTATC ACCAAATTAA GGGCAAAAGG
2101 CCAGGTTCTT ACAAAAACAA ACCCAAACAG CATTAAGAAG AAACAAAAGG
2151 ACCCTCAGGA CATCTGGAG GTGAAGGAAC GTGTAGAAAA AAACACCATG
2201 TTTTCTTCTC AAGCTGAGGA TGAATTGGAG CCTGCCAGGA AAAAAAGGAG
2251 AGAACAACTT GCCTATCTGG AATCTGAGGA ATTCAGAAA ATCTTAAAG
2301 CAAAATCAAA ACACACAGGC ATCCTGAAAG AGGCCGAGGC TGAGATGCAG
2351 GAGCGCTACT TTGAGCCACT GGTGAAAAAA GAACAAATGG AAGAAAAGAT
2401 GAGAAACATC AGAGAAGTGA AGTGCCGTGT CGTGACATGC AAGACGTGCG
2451 CCTATACCCA CTTCAAGCTG CTGGAGACCT CGCTCAGTGA GCAGCATGAA
2501 TACCACTGGC ATGATGGTGT GAAGAGGTTT TTCAAATGTC CCTGTGGAAA
2551 CAGAAGCATC TCCTTGAGCA GACTCCCGAA CAAGCACTGC AGTAACTGTG
2601 GCCTCTACAA ATGGGAACGG GACGGAATGC TAAAGTATG CCATTTGCGT
2651 ACTAATTTTT GACTCCTTTT AGTGACCCAT GCTAATAATG TGAACCATC
```

```

2701 TCCTATTA AAA ATATTTTCAT TTTTCTAGGA AAAGACTGGT CCAAAGATAG
2751 GAGGAGAAAC TCTGTTACCA AGAGGAGAAG AACATGCTAA ATTTCTGAAC
2801 AGCCTTAAAT AACCCGAAC TCAGACATTT TCCCACAGAC TTCTTGGCCT
2851 CCGTGACTG TGGAAAGCAA AGGATTGGCT GTGTATTGTC CATTGATTCC
2901 TGATTGACGC CGTCAAAAAC AAATGCTTGT TAAGCCCATTA AGCTTTGCCT
2951 GCTTACTTTC TGCCATTGGG TTGGTTTGAT ACCACATTTA ACATTGACAT
3001 TTAAGTGGAA AACCAAGTTA TCATTGTCTT TTCTAAGCTC AGTGTGGATG
3051 ATTGCATTAC TTCATTCACT GAAGTTTTTG CCCAAAAATT GGAAGGTAAA
3101 CAGAGAGCTA TGTTTCTGTA TCTTTTGGTT ATAGAGTGTT CACTTCTTTA
3151 TCATAACAAA ATTCTAGTGT TTATACGAAC ACCCAGAGGC AAAAGAATTT
3201 GGCTTAATTC TCACTCCAGG TAAGTAGCTT AACTTCTGGG CTTCAGTTTT
3251 CTCATCTGTA AAATCAGGAA GATTGGACTA AGTGATCCTG AAATGTATTT
3301 TTTAGCACTG GATTTCTACA AATAATAAAA CTTTCCCATC TAGATAATGA
3351 TGATCACATA GTCTTGATGT ACGGACATTA AAAGCCAGAT TTCTTCATTC
3401 AATTCTGTGA TCTCTGTTTT ACTCTTTGAA ATTGATCAAG CCACTGAATC
3451 ACTTTGCATT TCAGTTTATA TATAGAGAGA GAAAGAAGGC TGCTGCTCT
3501 TACATTATTG TGGAGCCCTG TGATAGAAAT ATGTAAATC TCATATTATT
3551 TTTTTTTTAA TTTTTTTTAT TTTTATGACA GGGTCTCACT ATGTCACCCCT
3601 GGCTGGAGTG CAGTAGTGCG ATCGCGGCAC ACTGCAGCCT TGGCTTCCCT
3651 GGGCTCAAGC AGTCCCTCCA CCTCAGTCTC CCAAAATAGCT AGGACTACAG
3701 GCGTGGCTGA CCAAGCCGAG CTAATTTTGT CATTTTGTGT AGAGATGGGG
3751 TTTTGCCATG TTGCTCAGGC TGGTCTCAAA CTCCTGAGCA CTAGCAATCC
3801 ACCCACCTCT GTTTCACAAA AAAAAAAAAA AATGAAAGGT CAACCCCTAT
3851 GCAAATTACC ACAGCAAAGG TTTCAATCAG GAGATTCTTC CATCTGGGCA
3901 ACCTGGTTTT CCAAATATCA TTTGACCTAA GTGAATGTTG ATACTAGCTA
3951 AAGATTGGGT AAATTGGTTG AATTATTGTA TTGAAGCTTG AGCTGTAGCT
4001 AAAAGTAATT TAGGTTTCCC CTAAGATGTT ATTATGTTAG GGACATAACA
4051 CTTTTGGGAG GTTGTGTGTT GAGATGGTTG ATTTAGGTTT TCAAAAGCTA
4101 GAAATAAAAT TTACATGCCT TAGATTTCAT AAAATTCTGC TCTAATTGGG
4151 TGGAAAGGTG TGTATCTAAC TTGTGTTTCT CTAAGGTTA TGTCCTAATA
4201 ACTATTCTTT TAGGAGTATA CTTCTACTTT ATAGAAGGTT GCTTTTCTTT
4251 TTAATTTTTT CTAACAAAGA AAAGAATAAA GTATTTATTA ATAAGAACCA
4301 GAAAGCACTT GAAACTGATG TTTTAAATGG CTCATTAGG GTAGATTAT
4351 TTATCTCATT AACTTAAAC AGCTATGTGT ATGAAATAGG TCACAACAGA
4401 ACTTGAACAC CAGGTGGTG TCTGAGCAAT CCCTTTCTTA TGGGAAAAAC
4451 AATGTTCTTG TTGAACAGA GGGTATCATT GCAGTCAGTA TTCACGTGTA
4501 TATTGTTATA TAAGTTGTAT AATATGCTTG TAAAGGCTGA GGGTGAGCTG
4551 TATCTGGATG CCTTTTACA ATTTGATTTT AACTTTTAAA ATAAATTTAA
4601 AACATAAAAA AAAAAAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 95 bp to 2659 bp; peptide length: 855
 Category: similarity to known protein
 Classification: Cell division

```

1 MDEEDNLSL LTALLEENES ALDCNSEENN FLTRENGEPD AFDELFDADG
51 DGESEYEEAD DGETGETRDE KENLATLFGD MEDLTDEEEV PASQSTENRV
101 LPAPAPRREK TNEELQEELR NLQEQMKALQ EQLKVTTIQK TASPRLQKS
151 PEKSPRPPLK ERRVQRIQES TCFSAEILDVP ALPRTKRVAR TPKSPPPDPK
201 SSSSRMTSAP SQPLQTI SRN KPSGITRGQI VGTGSSSET TQPICVEAFS
251 GLRLRRPRVS STEMNKKMTG RKLIRLSQIK EKMAREKLEE IDWVTFGVIL
301 KKVTPQSVNS GKTFSIWKLN DLRDLTQCVS LFLFGEVHKA LWKTEQGTVV
351 GILNANPMKP KDGSEEVCLS IDHPQKVLIM GEALDLGTCK AKKKNGEPT
401 QTVNLRDCEY QYHVQAQYK KLSAKRADLQ STFSGGRIPK KFARRGTSK
451 ERLCQDGFY GGVSSASYAA SIAAAVAPKK KIQTTLNVL VKGTNLIQ
501 TRQKLGIPQK SLSCSEEFKE LMDLPTCGAR NLKQLAKAS ASGIMGSPKP
551 AIKSISASAL LKQKQRMLE MRRRKSEIEQ KRFLQSSSEV ESPAVPSSSR
601 QPPAQPPRTG SEFPRLGAP ATMTPKLGRG VLEGDDVLFY DESPPRPKL
651 SALAEAKKLA AITKLRAKQ VLTKTNPNSI KKKQKDPQDI LEVKERVEKN
701 TMFSSQAED ELPARKRRE QLAYLESEEF QKILKAKSKH TGILKEAEAE
751 MQERYFEPLV KKEQMEEMR NIREVKCRVV TCKTCAYTHF KLETCVSEQ
801 HEYHWHDGVK RFFKPCGNR SISLRLPNK HCSNCGLYKW ERDGMKLVCH
851 LRTNF

```


BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2h15, frame 2

TREMBLNEW:SPBC1347_10 gene: "cdc23"; "SPBC1347.10"; product: "cell division cycle protein 23"; S.pombe chromosome II cosmid c1347., N = 2, Score = 284, P = 7e-21

PIR:S48384 DNA43 protein - yeast (Saccharomyces cerevisiae), N = 2, Score = 203, P = 7e-12

TREMBL:SCDNA52A_1 gene: "DNA52"; Saccharomyces cerevisiae DNA52 gene, complete cds., N = 2, Score = 201, P = 7.9e-12

TREMBLNEW:AC006234_6 gene: "F5H14.6"; Arabidopsis thaliana chromosome II BAC F5H14 genomic sequence, complete sequence., N = 2, Score = 211, P = 1.7e-15

PIR:S48384 DNA43 protein - yeast (Saccharomyces cerevisiae), N = 2, Score = 203, P = 7.2e-12

>TREMBLNEW:SPBC1347_10 gene: "cdc23"; "SPBC1347.10"; product: "cell division cycle protein 23"; S.pombe chromosome II cosmid c1347.
Length = 593

HSPs:

Score = 284 (42.6 bits), Expect = 7.0e-21, Sum P(2) = 7.0e-21
Identities = 97/383 (25%), Positives = 186/383 (48%)

Query: 109 EKTNEELQEELRNLEQMKALQEQLKVTTIQTASPARLQKSPEKSPRPPLKERRVQRIQ 168
E+ + +L+E + LQ Q+ +QE+ ++ + ++ AS + + PR P ++ RV +
Sbjct: 8 EENDLDLEE--KRLQRQLNEIQEKKRLRSAQKEASSENAEVI--QVPRSPQVRLTVS 63

Query: 169 ESTCFSAE----LDVPALPRTKRVARTPKPSPPDPKSSSSRMTSAPSQP-----LQTIS 218
+ + + L + K V+ P P PK R+ A +Q L+T+
Sbjct: 64 SPSKLPKPKRLILGIDKGTGKDVSLGKGRGPLPKPFHERLAEARNQERKRSCLKTKTMK 123

Query: 219 RNKPSGITRGQIVGTPGSSGETTQPI-C--VEAFSGLRLRRPRVSSTEMNKKMTGRKLIR 275
+N+ R + + G S E P+ C ++ +S + +S + + G ++
Sbjct: 124 KNRKQSFQRKRNILEDGKSEEEKFPMKCDEIDPYSRQAIVIRYISDEVAKENIGGNQVYL 183

Query: 276 LSQIKEKMAREKLE--EID-WVTFGVILKKV-TPQSVNSGKTFSIWKLNLDRLDLCVSL 331
+ Q+ + + K E E+D +V G++ T ++VN K + + L DL+ +C
Sbjct: 184 IHQLKLVRAPKFEAPEVDNYVVMGIVASNSGTRETVNGNK-YCMLTLTDLKWQLEC--- 239

Query: 332 FLFGVHKALWKTEQGTVVGILNANPMKPKDGS-EEVCLSIDHPQKVLI-MGEALDLGTC 389
FLFG+ + WK + GTV+ +LN +KPK+ L +D VL+ +G + LG C
Sbjct: 240 FLFGKAFERYWKIQSGTVIALLNPEVLKPKNPDIGRFSCLKDSEYDVLLEIGRSKHLGYC 299

Query: 390 KAKKKNGEPCQTQTVNLRDCEYCYHVQAQYKKLSAKRADLQSTFSGGRIPKKFARRGTS 449
+++K+GE C ++ R + C+YHV ++ + R + S+ + P+ ARR
Sbjct: 300 SSRKSGELCKHWDKRGDVCYHVDLAVQSRMSTRTEFASMATMHEPR--ARR--- 353

Query: 450 KERLCQDGF--YYGGVSSASASIAAAVAPKKKIQT 484
++R GF Y+ G ++ ++A + +QT
Sbjct: 354 EKRFRRGGFGYFAGEKYSIIPNAVAGLYDAEDAVQT 390

Score = 41 (6.2 bits), Expect = 7.0e-21, Sum P(2) = 7.0e-21
Identities = 12/43 (27%), Positives = 17/43 (39%)

Query: 453 LCQDGFYYGGVSSASASIAAAVAPKKKIQTTLNVLVVGKTN 495
L +D S AS A++ K + SN + GTN
Sbjct: 465 LSKDSEIDSSTKKPSVLASFNASIMNPKSSLPSFSNSAILGTN 507

Score = 40 (6.0 bits), Expect = 8.9e-21, Sum P(2) = 8.9e-21
Identities = 13/26 (50%), Positives = 18/26 (69%)

Query: 536 LAKASASGIMGSPKPAIKSISASALL 561
LA +AS IM +PK ++ S S SA+L
Sbjct: 481 LASFNAS-IM-NPKSSLPSFSNSAIL 504

Pedant information for DKFZphtes3_2h15, frame 2

Report for DKFZphtes3_2h15.2

(LENGTH) 855
 (MW) 96135.01
 (pI) 8.96
 (HOMOL) TREMBLNEW:SPBC1347_10 gene: "cdc23"; "SPBC1347.10"; product: "cell division
 cycle protein 23"; S.pombe chromosome II cosmid c1347. 5e-16
 (FUNCAT) 03.22 cell cycle control and mitosis [S. cerevisiae, YIL150c] 1e-11
 (FUNCAT) 03.16 dna synthesis and replication [S. cerevisiae, YIL150c] 1e-11
 (FUNCAT) 30.10 nuclear organization [S. cerevisiae, YIL150c] 1e-11
 (KW) Alpha_Beta
 (KW) LOW_COMPLEXITY 12.05 %
 (KW) COILED_COIL 4.21 %

SEQ MDEEEDNLSLLTALLEENESALDCNSEENNFLTRENGEPDAFDELFDADGDGESYTEEAD
 SEGxxxxx
 PRD cccchhhhhhhhhhhhhhhhhhhhhccccccccceeeccccccccceeeccccccccceeeec
 COILS

SEQ DGETGETRDEKENLATLFGDMEDLTDEEVPASQSTENRVLPAPAPRREKTNEELQEELR
 SEG xxxxxxxxxxxxxxxx.....xxxxxxxxxxxx
 PRD cccccccccchhhhhhhccccccccceeeccccccccccccccccccccchhhhhhhhhhhhh
 COILSCCCCCCCCCCCCC

SEQ NLQEQMKALQEQLKVTTIKQTASPARLQKSPEKSPRPPLKERRVQRIQUESTCFSAELDVP
 SEG xxxxx.....
 PRD hhhhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccceeecccccccccccccc
 COILS CCCCCCCCCCCCCCCCCCCC.....

SEQ ALPRTRKRVARTPKPSPPDPKSSSRMTSAPSQPLQTIISRNKPSGITRGQIVGTPGSSGET
 SEGxxxxxxxxxxxxxx
 PRD cccccceeeccccccccccccchhhhhhhccccchhhhhhhccccceeecccccccccc
 COILS

SEQ TQPICVEAFSGLRLRRPRVSTEMNKKMTGRKLIRLSQIKEKMAKLEEDWVTFGVIL
 SEG
 PRD cccccccccchhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccceeeeee
 COILS

SEQ KKVTPQSVNSGKTFSIWKLNDLRDLTQCVSLFLFGEVHKALWKTEQGTVVGILNANPMKP
 SEG
 PRD cccccccccccccceeeccccchhhhhhhheeeccccchhhhhhhhhccccceeecccccccc
 COILS

SEQ KDGSEEVCLSIDHPQKVLIMGEALDLGTCKAKKKNGEPTQTVNLRDCEYCYHVQAQYK
 SEG
 PRD cccccceeeccccccccceccccccccccccccccccccceeeccccccccchhhhhhhhh
 COILS

SEQ KLSAKRADLQSTFSGGRIPKKFARRGTSILKERLCQDGFYGGVSSASAYAASIAAAVAPKK
 SEGxxxxxxxxxxxxxxxxxxxxxx
 PRD hhhhhhhhhhhhhccccccccccccccccchhhhhhhccccccccccccchhhhhhhhhhhhhccch
 COILS

SEQ KIQTTLNVLVVGKTNLIQETRQKLGIPQKSLSCSEEFKELMDLPTCGARNLKQHLAKAS
 SEG
 PRD hhhhhhhheeeccccceeehhhhhhccccccccchhhhhhhhhccccccccchhhhhhhhhhh
 COILS

SEQ ASGIMGSPKPAIKSISASALLKQKQRMLEMRKSEEIQKRFLOSSEVESPAVPSSSR
 SEGxxxxxxxxxxxxxxxxxxxxxx
 PRD hhccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccccccc
 COILS

SEQ QPPAQPPRTGSEFPRLEGAPATMTPKLGRGVLEGDDVLFYDESPPPRKLALAEAKKLA
 SEG xxxxxxxx.....xxxxxxxxxxxxxx
 PRD cccccccccccccccccccccccccccccccccceeeccccccccchhhhhhhhhhhhhhh
 COILS

SEQ AITKLRAKGQVLTKTNPSIKKKQKDPQDILEVKERVEKNTMFSSQAEEDELEPARKKRRE
 SEG xxxxx.....
 PRD hhhhhhhhhheeeccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
 COILS

SEQ QLAYLESEEFQKILKAKSKHTGILKEAEAEQERYFEPLVKKEQMEKMRNIREVKCRVV
 SEG
 PRD hhhhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhheee
 COILS

SEQ TCKTCAYTHFKLLETVCSEQHEYHWHGDGVRFFKPCGNRSISLDRLPNKHCSNCGLYKW

```

SEG      .....
PRD      eeeccceeeeecccccceccccccccceeeeeccccccccccccccccccccccccccceec
COILS    .....

```

```

SEQ      ERDGMLKVCHLRTNF
SEG      .....
PRD      cccccccccccccccc
COILS    .....

```

(No Prosite data available for DKFZphtes3_2h15.2)

(No Pfam data available for DKFZphtes3_2h15.2)

DKFZphtes3_2i5

group: testes derived

DKFZphtes3 2i5 encodes a novel 151 amino acid protein with weak similarity to. C.elegans
cosmid F20D12.3

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific
genes.

similarity to C.elegans F20D12.3

many ATGs in front of the start of the ORF,
unspliced intron in 5' region?

Sequenced by EMBL

Locus: unknown

Insert length: 2142 bp

Poly A stretch at pos. 2121, polyadenylation signal at pos. 2102

```
1 GCAGTAAATA TGATATGAAA GAATTCTCTA ACTTGGGGGT GGCTTGTAAC
51 CTGTAATAAA AATATTGCTA AAATACCTTC TCTCACTTTG AAAAAGCATC
101 TGAGCAATCC TCAGTTATTG GTGAATTCTT ACCAGTGTTC AATTCCTCTC
151 TTTCCGTTAT GGTCTTAGTG TGGTTGTCCT GGTGTAGTAT TTCAAGAGGA
201 ACCTGCAGCA AGATGAAAAG AGAGTGGGAC TTGGAGCTAA GAACGTTTTT
251 GGCTTTAAGT GCTACGTTAA CTCATTAAAT TCTTAGTGAT CTTGGGGGAA
301 TCCCTTCACC AGTGTGAGCC TCAGTTTCTT TATCTAATAA GTAAGGATAA
351 TCTTACCCAC CTTATTGCGG GGGCCCGAGG ATTACATGAT TGGTGTAAACA
401 GTAGCACCTT GTACATTTGA AAGGACTAAT ACCAGTGGAC TTAAACCTTG
451 TGGTGTACTT GGAATTCTTG GTGGGACTTT TTAATCATGT AGATTCTCAG
501 GCCCTGTCCT GGCCTGTGGA ACCACAGACT CTATAGGTGG GCCCTTCCAG
551 AAGGCCTCAT GGGTGGTTCT CATGTGGAAC CTGTGTTGCA AGCCACTGCA
601 TGGTGTACTT GCTATTAACA TTAAACTTAA TATTTTCCTT ATTGTGTGGA
651 TATATCTGTG GTGTTTGCCC ATGTATACTT CATTTTACAT TTCTTAAAGA
701 ATAGAATGGA ATGGTTTTAA GCACGCTACA TTGTCCAGGT TATACCCACA
751 GAAGAGCTGT TGTGTAACAG AATCAGCATC ATACCTGAAT CATTTGTACA
801 TTGCATATAA GACTATGTCT AAGTAGAAGA TGCTATGAAA TCATGTCTGC
851 TGTGGGGCCA GGCATAATTA TGAATGTTAC TTAAGAGCAT AGGTGAGGTG
901 AGAAAAAGGA ATGTGACTAG TGTTTTAGTA TTTTCTTGTT GTGGGATGAA
951 GTATAATTCT TTTTTTTTTT TCTCAACAAA GCAGTAAAAC TAGAAAAGAA
1001 GAGAACTCTT CCCTCAAGAA TGGCTGTACC TTCATATCTA GAGGCACATT
1051 AAAAAAAGA ACGTCTGTAC CTTAAAAATG GAGGTCAATT CATTTGTGTT
1101 ATTTTCAAGG TTGTTGTATG GCTCGGTCAG AACTTTCTGT TACCAGAAGA
1151 CACTCACATT CAGAATGCTC CATTTCAAGT GTGTTTCACA TCTTTACGGA
1201 ATGGCGGCCA CCTGCATATA AAAATAAAAC TTAGTGGAGA GATCACTATA
1251 AATACTGATG ATATTGATTT GGCTGGTGAT ATCATCCAGT CAATGGCATC
1301 ATTTTGTGCT ATTGAAGACC TTCAAGTAGA AGCGGATTTT CCTGTCTATT
1351 TTGAGGAATT ACGAAAGGTG CTAGTTAAGG TGGATGAATA TCATTCAGTG
1401 CATCAGAAGC TCAGTGCTGA TATGGCTGAT CATTTCTAAT TGATCCGAAG
1451 TTTGCTGGTC GGAGCTGAGG ATGCTCGTCT GATGAGGGAC ATGAAAACAA
1501 TGAAGAGTCG TTATATGGAA CTCTATGACC TTAATAGAGA CTTGCTAAAT
1551 GGATATAAAA TTCGCTGTAA CAATCACACA GAGCTGTTGG GAAACCTCAA
1601 AGCAGTAAAT CAAGCAATTC AAAGAGCAGG TCGTCTGCGG GTTGGAAAAA
1651 CAAAGAACCA GGTGATCACT GCTTGTGCGG ATGCAATTCG AAGCAATAAC
1701 ATCAACACAC TGTTCAAAAT CATGCGAGTG GGGACAGCTT CTTCTAGGT
1751 GAGGAAAAATA CAGGTCATGA AGTTCCTGGC AAAGATTTTC TGTAAAAAAC
1801 CTATGCTGGT TTGCTTTGGA TCACACCTGT GTGAACCCCG GGTGCTAAGA
1851 ATGAAAATAA CTTTGGTGAG TTGTACAAAT TAAAGACAAA GAACTACATG
1901 TGAAGATAGA CTTGCTTTCT ATTTTAAAT CAGTAGTAGT ACTGTTGCTG
1951 AATAATACTA GGTTTTTATG GAATAGGATG AATGCTTTTG AAGTATTAGG
2001 GCTTCAGAGT CCAATTTTGC TTATTTATGG TATATAAATA CATATTTTTT
2051 TCTTGAAATT GCAATTGAGT TTGTACTTTT CAAATAGATT ATCTACTTTT
2101 TCATTAATAA GTAAAGATGT TAAAAAATAA AAAAAAATAA AA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 1293 bp to 1745 bp; peptide length: 151
Category: similarity to unknown protein
Classification: no clue

1 MASFFAIEDL QVEADFPVYF EELRKVLVKV DEYHSVHQKL SADMADHSNL
51 IRSLLVGAED ARLMRDMKTM KSRYMELYDL NRDLLNGYKI RCNNHTELLG
101 NLKAVNQAIQ RAGRLRVGKP KNQVITACRD AIRSNNINTL FKIMRVGTAS
151 S

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2i5, frame 3

TREMBL:CEF20D12_1 gene: "F20D12.3"; Caenorhabditis elegans cosmid
F20D12., N = 1, Score = 173, P = 4.5e-12

>TREMBL:CEF20D12_1 gene: "F20D12.3"; Caenorhabditis elegans cosmid F20D12.
Length = 699

HSPs:

Score = 173 (26.0 bits), Expect = 4.5e-12, P = 4.5e-12
Identities = 33/130 (25%), Positives = 72/130 (55%)

Query: 20 FEELRKVLVKVDEYHSVHQKLSADMADHSNLIRSLLVGAEDARLMRDMKTMKSRYMELYD 79
F+E ++L ++D V +L+A++ + ++ +++ AED+ + ++ + Y+ L
Sbjct: 569 FKEADEILEEIDPMTEVRDLTAELQERQAAVKEIIIRAEDSIAIDNIPDARKFYIRLKA 628
Query: 80 LNRDLLNGYKIRCNNHTELLGNLKAVNQAIQRAGRLRVGKPKNQVITACRDAIRSNNINT 139
+ ++R NN + +L+ +N+ I+ RLRVG+P Q++ +CR AI +N
Sbjct: 629 NDAARQAQLRWNNQERCVKSLRRLNKIIENCRLRVGEPGRQIVVSCRSIAIADDNKQI 688
Query: 140 LFKIMRVGTA 149
+ KI++ G +
Sbjct: 689 ITKILQYGAS 698

Pedant information for DKFZphtes3_2i5, frame 3

Report for DKFZphtes3_2i5.3

[LENGTH] 151
[MW] 17304.07
[pI] 9.33
[HOMOL] TREMBL:CEF20D12_1 gene: "F20D12.3"; Caenorhabditis elegans cosmid F20D12. 2e-12
[KW] Alpha_Beta

SEQ MASFFAIEDLQVEADFPVYFEELRKVLVKVDEYHSVHQKLSADMADHSNLIRSLLVGAED
PRD cceeeehhhhhccccchhhhhhhhhhhccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
SEQ ARLMRDMKTMKSRYMELYDLNRDLLNGYKIRCNNHTELLGNLKAVNQAIQRAGRLRVGKP
PRD hhhhccchhhhhheeeccchhhhhheeeccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
SEQ KNQVITACRDAIRSNNINTLKFIMRVGTASS
PRD cceeeehhhhhccccceeeccceeeccccc

(No Prosite data available for DKFZphtes3_2i5.3)

(No Pfam data available for DKFZphtes3_2i5.3)

DKFZphtes3_2119

group: testes derived

DKFZphtes3_2119 encodes a novel 166 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, no EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 1079 bp

Poly A stretch at pos. 1053, polyadenylation signal at pos. 1038

```
1 CCACAGGACA CACTGTCCCC AGGGCACAGA CACCCTGGGC TTTGGTTGGG
51 TCTTGGCCTC CAGGTAGGGC CCTGTTGGGC AGCGGGCAGC AACTCCTGAG
101 ACACACTACTGT GATTCTTGGT GGTGGCTGTG GTAAAAAACC TGCAGGGCTA
151 GAGTTTGGGG TGAGATTCAG CAGTAACTGT GGCCTCTCCT AGTGACAGTA
201 TGCTACTCCC ACTCCCAGCA CGCATGCCCA CAGGCCACGG CCTCCACATC
251 ACAAAACCCC CACCAAGTTG CCCATCTATG GAGCAGCTCC CATAACGGCAG
301 GGTGAGGCTC TTACCTCCAC CTCCAGGGCA CAGACAGGGG GAGCTCTGTC
351 TCACTGTAAG GCAATGAGGA GAGTTGAGGG CCCAGACCAG GCTAGGGGCC
401 ATCCCCTTTC CCGAGCAGGC CTCAGGGAAG GACCAGCCCC ATTCCCATCT
451 GACCTAGGTC TTAGCCCAGG AGCCTGCATA GGGAGAAAG GACAGACAGG
501 GCCTCCTTAC TGGCTGACAC TCAGGAGGGG CTGGGGCAAG AGAGCAGAGG
551 GAGCGCAGGG CCAGGCAGGG GCTGCTGAGG ATCCATGGGA GCTCAGGGTG
601 CACAAGGGGG CTGCCCTTCC TGGGCTGCAG GCAGCATCCC TATGGGAGCT
651 GAGAAAGTCC AATCCTGAGA TGGGACAGTG CTGCCCAGGG GTGTGTGGCT
701 GGGCCCTGAC AACAGTCTCC CCAAAGTGA CCACATCACC AGGCTCAGTT
751 CCAGGAAGGC TGAGAAGTGC CCAGTACACT GAGGATGCAC CTCAGTTACA
801 TAAATATAAT GAAACTGGAG TACTAACGTA CAGTTTAAAG GTTATAGTTA
851 CTATTTTAT ATGATATACT AGTAATTTT GAATAGGTA AACTTTAGGT
901 GTTTTGACAC CAAAGAAAA CTACATGAGT TCATGCATGT GTTAAATTGC
951 TTTACTGTAG TAATCATTTA CATGTATATG TATATATGAA TATAATTATG
1001 GGCTCATTA ATTTAAATAT TATAAATAGG TGACAAAGAA TAAAGTTAAC
1051 TGGAAAAAA AAAAAA AAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 364 bp to 861 bp; peptide length: 166
Category: putative protein
Classification: no clue

```
1 MRRVEGPDQA RGHPLSRAGL REGPAPFSPD LGLSPGACIG KKGQTGPPYW
51 LTLRRGWGKR AEGAQGOAGA AEDPWELRVH KGAALPGLQA ASLWELRKS
101 PEMGQCCPGV CGWALTTVSP KVTTSFGSVP GRLRSAQYTE DAPQLHKINE
151 TGVLTYSKLV IVTIFI
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2119, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_2119, frame 1

Report for DKFZphtes3_2119.1

{LENGTH} 166
{MW} 17691.35
{pI} 9.54
{KW} All_Beta
{KW} LOW_COMPLEXITY 7.23 %

SEQ MRRVEGPDQARGHPLSRAGLREGPAPFPSPDLGLSPGACIGKKGQTGPPYWLTLRRGWGKR
SEG
PRD ccc

SEQ AEGAQQGAGAAEDPWELRVHKGALPGLQAASLWELRKSNPENMGQCCPGVCGWALTTVSP
SEG xxx
PRD ccc

SEQ KVTTSPPGSPGRLRSAQYTEDAPQLHKINETGVLTYSLKVIVTIFI
SEG
PRD ccc

(No Prosite data available for DKFZphtes3_2119.1)

(No Pfam data available for DKFZphtes3_2119.1)

DKFZphtes3_2m18

group: nucleic acid management

DKFZphtes3_2m18 encodes a novel amino acid protein, with similarity to mouse Dhml.

The protein seems to play a role in nucleotide metabolism, RNA metabolism, but also in DNA repair and cell cycle. The yeast homologue is a DNA strand exchange protein required for sporulation and homologous recombination.

The novel protein can find application as multifunctional nuclease / exoribonuclease.

nearly identical to mouse Dhml

complete cDNA, complete cds, start at Bp 42, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 3022 bp

Poly A stretch at pos. 3004, polyadenylation signal at pos. 2981

```

1 CTCGTCAGCC GGTCCGCCGC CGCCTCCAGC CGTGTGCCGC TATGGGAGTC
51 CCGGCGTTCT TCCGCTGGCT CAGCCGCAAG TACCCGTCCA TCATAGTCAA
101 CTGCGTGGAA GAGAAGCCAA AAGAATGCAA TGGTGTAAG ATCCAGTTG
151 ATGCCAGTAA ACCTAATCCA AATGATGTGG AGTTTGATAA TCTGTATTG
201 GATATGAATG GAATCATCCA TCCCTGTACT CATCCTGAAG ACAAACCAGC
251 ACCAAAAAAT GAAGATGAAA TGATGGTTGC AATTTTTGAG TACATTGACA
301 GACTTTTCAG TATTGTAAGA CCAAGAAGAC TTCTCTACAT GGCAATAGAT
351 GGAGTGGCAC CACGTGCTAA AATGAACCAG CAGCGTTCAA GGAGGTTGAG
401 GGCATCAAAA GAAGGAATGG AAGCAGCAGT CGAGAAGCAG CGAGTCAGGG
451 AAGAAATATT GGCAAAAGGT GGCTTTCTTC CTCCAGAAGA AATAAAGAA
501 AGATTTGACA GCAACTGTAT TACACCAGGA ACTGAATTCA TGGACAATCT
551 TGCTAAATGC CTTGCTATT ACATAGCTGA TCGTTTAAAT AATGACCCGT
601 GGTGGAAAAA TTGACAGTT ATTTTATCTG ATGCTAGTGC TCCTGGTGAA
651 GGAGAACATA AAATCATGGA TTACATTAGA AGGCAAGAG CCCAGCCTAA
701 CCATGACCCA AATACTCATC ATTGTTTATG TGGAGCAGAT GCTGATCTCA
751 TTATGCTTGG CCTTGCCACA CATGAACCGA ACTTTACCAT TATTAGAGAA
801 GAATTCAAAC CAAACAAGCC CAAACCATGT GGTCTTTGTA ATCAGTTTGG
851 ACATGAGGTC AAAGATTGTG AAGGTTTGCC AAGAGAAAAG AAGGGAAAGC
901 ATGATGAATC TGCCGATAGT CTTCTTGTG CAGAAGGAGA GTTTATCTTC
951 CTTCCGGCTT ATGTTCTTCG TGAGTATTTG GAAAGAGAAC TCACAATGGC
1001 CAGCCTACCA TTCACATTG ATGTTGAGAG GAGCATTGAT GACTGGGTTT
1051 TCATGTGCTT CTTTGTGGGA AATGACTTCC TCCCTCATT GCCATCGTTA
1101 GAGATTAGGG AAAATGCAAT TGACCGTTTG GTTAACATAT ACAAAAATGT
1151 GGTACACAAA ACTGGGGGTT ACCTTACAGA AAGTGGTTAT GTCAATCTGC
1201 AAAGAGTACA GATGATCATG TTAGCAGTTG GTGAAGTTGA GGATAGCATT
1251 TTTAAAAAGA GAAAGGATGA TGAGGACAGT TTTAGAAGAC GACAGAAAGA
1301 AAAAAAGAA AGAATGAAGA GAGATCAACC AGCTTTCACCT CAGTAGGGAA
1351 TATTAATCC TCATGCCTTG GGTTCAGAA ATTCAACAGG TTCTCAAGTA
1401 GCCAGTAATC CGAGACAAGC AGCCTATGAA ATGAGGATGC AGAATAACTC
1451 TAGTCCTTCG ATATCTCCTA ATACGAGTTT CACATCTGAT GGCTCCCCGT
1501 CTCCATTAGG AGGAATTAAG CGAAAAGCAG AAGACAGTGA CAGTGAACCT
1551 GAGCCAGAGG ATAATGTCAG GTTATGGGAA GCTGGCTGGA AGCAGCGGTA
1601 CTACAAGAAC AAATTTGATG TGGATGCAGC TGATGAGAAA TTCCGTCGGA
1651 AAGTTGTGCA GTCGTACGTT GAAGGACTTT GCTGGGTTCT TAGATATTAT
1701 TACCAGGGCT GTGCTTCCTG GAAGTGGTAT TATCCATTTC ATTATGCACC
1751 ATTTGCTTCA GACTTTGAAG GCATTGCAGA CATGCCATCT GATTTTGAGA
1801 AGGGTACGAA ACCGTTTAAA CCACTAGAAC AACTTATGGG GGTATTTCCT
1851 GCTGCAAGTG GTAATTTTCT ACCTCCATCA TGGCGGAAGC TCATGAGTGA
1901 TCCTGATTCT AGTATAATTG ACTTCTATCC TGAAGATTTT GCTATTGATT
1951 TGAATGGGAA GAAATATGCA TGGCAAGGTG TTGCTCTCTT GCCATTCTGT
2001 GATGAGCGAA GGCTACGAGC TGCCCTAGAA GAGGTATACC CAGACCTCAC
2051 TCCAGAAGAG ACCAGAAGAA ACAGCCTTGG AGGTGATGTC TTATTTGTGG
2101 GGAACATCA CCCACTCCAT GACTTCATTT TAGAGCTGTA CCAGACAGGT
2151 TCCACAGAGC CAGTGGAGGT ACCCCCTGAA CTATGTCATG GGATTCAAGG
2201 AAAGTTTTCT TTGGATGAAG AAGCCATTCT TCCAGATCAA ATAGATGTTT
2251 CTCTGTTCCT TATGTTAAGG GATCTGACAC AGAACACTGT AGTCAGTATT
2301 AATTTTAAAG ACCACAGTT TGCTGAAGAT TACATTTTAA AAGCTGTAAT
2351 GCTTCCAGGA GCAAGAAAGC CAGCAGCAGT ACTGAAACCT AGTGACTGGG
2401 AAAAATCCAG CAATGGACGG CAGTGGAAAG CTCAGCTTGG CTTTAAACCGT
2451 GACCGGAGGC CTGTGCACCT GGATCAGGCA GCCTTCAGGA CTTTGGGCCA
2501 TGTGATGCCA AGAGGCTCAG GAACTGGCAT TTACAGCAAT GCTGCCACCA
2551 CACCTGTGAC TTACAGGGA AACTTATACA GGCCGCTTTT GAGAGGACAA
2601 GCCCAGATTC CAAAATTAT GTCAAAATAT AGGCCCCAGG ATTCCTGGCG
2651 AGGTCTCTCT CCCCTTTTCC AGCAGCAAAG GTTTGACAGA GGCGTTGGGG

```



```

2701 CTGAACCTCT GCTCCCATGG AACCGGATGC TGCAAACCCA GAATGCAGCC
2751 TTCCAGCCAA ACCAGTACCA GATGCTAGCT GGGCCTGGTG GGTATCCACC
2801 CAGACGAGAT GATCGTGGAG GGAGACAGGG ATATCCCAGA GAAGGAAGGA
2851 AATACCTTTT GCCACCACCC TCAGGAAGAT ACAATTGGAA TTAAGCTTTT
2901 GTAAAGCTTT CCCAAATCCT TTCATCATTG TACAGTTTGA TGCTATTTGT
2951 GGAAAGATTG CCTTCTCAAG TAGTAGTTTT TAATAAACT ACAGTACTTT
3001 GTGTAAAAAA AAAAAAAAAA AA

```

BLAST Results

No BLAST result

Medline entries

95192042:

Characterization of cDNA encoding mouse homolog of fission yeast dhpl+ gene: structural and functional conservation.

97361754:

Cloning and characterization of mouse Dhml cDNA, a functional homolog of budding yeast SEPl.

Peptide information for frame 3

ORF from 42 bp to 2891 bp; peptide length: 950
Category: strong similarity to known protein

```

1 MGVPFAFFRWL SRKYPSTIIVN CVEEKPKECN GVKIPVDASK PNPNDVEFDN
51 LYLDMMNGIIH PCTHPEDKPA PKNEDEMMVA IFEYIDRLFS IVRPRRLLYM
101 AIDGVAPRAK MNQQRSSRRFR ASKEGMEAAV EKQVRREEIL AKGGFLPPEE
151 IKERFDSNCI TPGTEFMDNL AKCLRYIIAD RLNDPBGWKN LTVILSDASA
201 PGEGEHKIMD YIRRQRAQPN HDPNTHHCLC GADADLIMLG LATHEPNFTI
251 IREEFKPNKP KPCGLCNQFG HEVKDCEGLP REKKGKHDEL ADSLPCAEGE
301 FIFLRNLVLR EYLERELTMA SLPFTFDVER SIDDWVFMCF FVGNDFLPHL
351 PSLEIRENAI DRLVNIYKNV VHKTGGYLTE SGVNLQVRVQ MIMLAVGEVE
401 DSIFKKRKDD EDSFRRRQKE KRKRMRKRDQP AFTPSGILTP HALGSRNSPG
451 SQVASNPQQA AYEMRMQNNS SPSISPNTSF TSDGSPSPLG GIKRKAEDSD
501 SEPEPEDNVR LWEAGWKQRY YKNKFDVDAA DEKFRKRVVQ SYVEGLCWVL
551 RYYYQGCASW KWYYPFHYAP FASDFEGIAD MPSDFEKGTK PFKPLEQLMG
601 VFPAASGNFL PPSWRKLMSD PDSSIIDFYP EDFAIDLNGK KYAWQGVALL
651 PFVDERRLRA ALEEVPDLT PEETRNSLG GDVLFVGKHH PLHDFILELY
701 QTGSTPEVEV PPELCHGIQG KFSLDDEAIL PDQIVCSPVP MLRDLTQNTV
751 VSINFKDPQF AEDYIFKAVM LPGARKPAAV LKPSDWEKSS NGRQWKPOLG
801 FNRDRRPVHL DQAAFTLGH VMPRGSGTGI YSNAAPPPVT YQGNLYRPLL
851 RQQAQIPKLM SNMRPQDSWR GPPPLFQQQR FDRGVGAEP L PWNRLQQTQ
901 NAAFPQNYQ MLAGPGGYPP RRDRGRGRQG YPREGRKYPL PPSGRYNWN

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_2m18, frame 3

PIR:I49635 mouse Dhml protein - mouse, N = 1, Score = 4765, P = 0

PIR:S43891 dhpl protein - fission yeast (Schizosaccharomyces pombe), N = 3, Score = 1172, P = 2e-197

PIR:S20126 exoribonuclease RAT1 (EC 3.1.11.-) - yeast (Saccharomyces cerevisiae), N = 2, Score = 1146, P = 3.8e-175

PIR:S72531 exonuclease II - fission yeast (Schizosaccharomyces pombe), N = 4, Score = 622, P = 4.2e-125

>PIR:I49635 mouse Dhml protein - mouse
Length = 947

HSPs:

Score = 4765 (714.9 bits), Expect = 0.0e+00, P = 0.0e+00
Identities = 884/930 (95%), Positives = 895/930 (96%)

```

Query:      1  MGVPAFFRWLSRKYPSSIIVNCVEEKPKECNGVKIPVDASKPNPNDVEFDNLYLDMNGIIH 60
             MGVPAFFRWLSRKYPSSIIVNCVEEKPKECNGVKIPVDASKPNPNDVEFDNLYLDMNGIIH
Sbjct:      1  MGVPAFFRWLSRKYPSSIIVNCVEEKPKECNGVKIPVDASKPNPNDVEFDNLYLDMNGIIH 60

Query:     61  PCTHPEDKPAPKNEDEMMVAIFEYIDRLF+IVRPRRLLYMAIDGVAPRAKMNQQRSSRRFR 120
             PCTHPEDKPAPKNEDEMMVAIFEYIDRLF+IVRPRRLLYMAIDGVAPRAKMNQQRSSRRFR
Sbjct:     61  PCTHPEDKPAPKNEDEMMVAIFEYIDRLF+IVRPRRLLYMAIDGVAPRAKMNQQRSSRRFR 120

Query:    121  ASKEGMEAAVEKQVRVEEILAKGGFLPPEEIKERFDSNCITPGTEFMDNLAKCLRYIIAD 180
             A K GMEAAVEKQVRVEEILAKGGFLPPEEIKERFDSNCITPGTEFMDNLAKCLRYIIAD
Sbjct:    121  AIKGGMEAAVEKQVRVEEILAKGGFLPPEEIKERFDSNCITPGTEFMDNLAKCLRYIIAD 180

Query:    181  RLNNDPGWKNLTVILSDASAPGEGEHKIMDYIRRQRAQPNHDPNTHHCLCGADADLIMLG 240
             RLNNDPGWKNLTVILSDASAPGEGEHKIMDYIRRQRAQPN DPNTHHCLCGADADLIMLG
Sbjct:    181  RLNNDPGWKNLTVILSDASAPGEGEHKIMDYIRRQRAQPNQDPNTHHCLCGADADLIMLG 240

Query:    241  LATHEPNFTIIREEFKPNKPKPCGLCNQFGHEVKDCEGLPREKKGKHDELADSLPCAEGE 300
             LATHEPNFTIIREEFKPNKPKPC LCNQFGHEVKDCEGLPREKKGKHDELADSLPCAEGE
Sbjct:    241  LATHEPNFTIIREEFKPNKPKPCALCNQFGHEVKDCEGLPREKKGKHDELADSLPCAEGE 300

Query:    301  FIFLRNLVLEYLELERELTMASLPFTFDVERSIDDWVFMCFVVGNDFLPHLPSLEIRENAI 360
             FIFLRNLVLEYLELERELTMASLPF FDVERS DDW FMCFFVGNDFLPHLPSLEIRE AI
Sbjct:    301  FIFLRNLVLEYLELERELTMASLPFFFDVERSNDDEWFMCFFVGNDFLPHLPSLEIREGAI 360

Query:    361  DRLVNIYKNVVHKTGGYLTESGYVNLQVRQMIMLAVGEVEDSIFKKRKDDDEDSFRRRQKE 420
             DRLVNIYKNVVHKTGGYLTESGYVNLQVRQMIMLAVGEVEDSIFKKRKDDDEDSFRRRQKE
Sbjct:    361  DRLVNIYKNVVHKTGGYLTESGYVNLQVRQMIMLAVGEVEDSIFKKRKDDDEDSFRRRQKE 420

Query:    421  KRKRMRDQPAFTPSGILTPHALGSRNSPGSQVASNPRQAAYEMRMQNNSSPSISPNTSF 480
             KRKRMRDQPAFTPSGILTPHALGSRNSPG QVASNPRQAAYEMRMQ NSSPSISPNTSF
Sbjct:    421  KRKRMRDQPAFTPSGILTPHALGSRNSPGCQVASNPRQAAYEMRMQNNSSPSISPNTSF 480

Query:    481  TSDGSPSPLGGIKRKAEDSDSEPEPEDNVRLWEAGWKQRYKKNFDDVADEKFRKRVVQ 540
             SDGSPSPLGGI+RKAEDSDSEPEPEDNVRLWEAGWKQRYKKNFDDVADEKFRKRVVQ
Sbjct:    481  ASDGSPSPLGGIKRKAEDSDSEPEPEDNVRLWEAGWKQRYKKNFDDVADEKFRKRVVQ 540

Query:    541  SYVEGLCWVLRYYYQGCASWKWYYPFHYAPFASDFEGIADMSPDFEKGTKPFKPLEQLMG 600
             SYVEGLCWVLRYYYQGCASWKW YPFHYAPFASDFEGIADM S+FEKGTKPFKPLEQLMG
Sbjct:    541  SYVEGLCWVLRYYYQGCASWKWLYYPFHYAPFASDFEGIADMSSEFEKGTKPFKPLEQLMG 600

Query:    601  VFPAASGNFLPPSWRKLMSDPDSSIIDFYPEDFAIDLNGKKYAWQGVALLPFVDERRLRA 660
             VFPAASGNFLPP+WRKLMSDPDSSIIDFYPEDFAIDLNGKKYAWQGVALLPFVDERRLRA
Sbjct:    601  VFPAASGNFLPPTWRKLMSDPDSSIIDFYPEDFAIDLNGKKYAWQGVALLPFVDERRLRA 660

Query:    661  ALEEVYPDLTPEETRNSLGGDVLVFGKHHPLHDFILELYQTGSTPEVVPPELCHGIQG 720
             ALEEVYPDLTPEE RRNSLGGDVLVFGK HPL DFILELYQTGSTPEV+VPPPELCHGIQG
Sbjct:    661  ALEEVYPDLTPEENRRNSLGGDVLVFGKLHPLRDFILELYQTGSTPEVDVPPPELCHGIQG 720

Query:    721  KFSLDEEAILPDQIVCSPVPMRLDLTQNTVVSINFKDPQFAEDYIFKAVMLPGARKPAAV 780
             FSLDEEAILPDQ VCSVPVPMRLDLTQNT VSINFKDPQFAEDY+FKA MLPGARKPA V
Sbjct:    721  TFSLDEEAILPDQTVCSVPVPMRLDLTQNTAVSINFKDPQFAEDYVFAAMLPGARKPATV 780

Query:    781  LKPSDWEKSSNGRQWKPLGPNRRPVLHDQAAFRTLGHVMPRGSGTGIYSNAAPPPVT 840
             LKP DWEKSSNGRQWKPLGPNRRPVLHDQAAFRTLGHV PRSGT +Y+N A P
Sbjct:    781  LKPGDWEKSSNGRQWKPLGPNRRPVLHDQAAFRTLGHVTPRGSGTSVYTNALLPAN 840

Query:    841  YQGNLYRPLLRGQAQIPKLMSNMRPQDSWRGPPPLFQQQRFDRGVGAEPPLPWNRMQLTQ 900
             YQGN YRPLLRGQAQIPKLMSNMRP+DSWRGPPPLFQQ RF+R VGAEPPLPWNRM+Q Q
Sbjct:    841  YQGNLYRPLLRGQAQIPKLMSNMRPKDSWRGPPPLFQQHRRFERSVGAEPPLPWNRMILNQ 900

Query:    901  NAAFQPNQYQMLAGPGGYPPRRDD-RGGRQ 929
             NAAFQPNQYQML GPGGYPPRRDD RGGRQ
Sbjct:    901  NAAFQPNQYQMLGGPGGYPPRRDDHRGGRQ 930

```

Pedant information for DKFZphtes3_2m18, frame 3

Report for DKFZphtes3_2m18.3

```

[LENGTH]      950
[MW]           108582.68
[pI]           7.26
[HOMOL]        PIR:I49635 mouse Dhml protein - mouse 0.0
[FUNCAT]       08.01 nuclear transport [S. cerevisiae, YOR048c] 1e-123
[FUNCAT]       04.01.04 rRNA processing [S. cerevisiae, YOR048c] 1e-123

```

```

SEQ      MGVPAFFRWLSRKYPSSIIVNCVVEEKPKCEKNGKIPVDASKPNPNDVEFDNLNLYDMNGIHH
SEG
PRD      cccchhhhhhhhhhhccceeeeecccccccccccccccccccccccccccccccccccccccccc
MEM      .....

SEQ      PCTHPEDKPAPKNEDEMMVAIFEYIDRLFSIVRPRRLLYMAIDGVAPRAKMNQORSRRFR
SEG
PRD      cccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccceeeeeccccchhhhhhhhhhhhhhh
MEM      .....

SEQ      ASKEGMEAAVEKQVRVEEILAKGGFLPPEEIKERFDSNCITPGTEFMDNLAKCLRYIIA
SEG
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM      .....

SEQ      RLNNDPGWKNLTVILSDASAPGEGEHKIMDYIRRQRAQPNHDPNTHHCLCGADADLIMLG
SEG
PRD      hccccccccceeeeeccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM      .....

SEQ      LATHEPNFTIIREEFKPNKPKPCGLCNQFGHEVKDCEGLPREKKGKHDELADSLPCAEGE
SEG
PRD      cccccccccccccccccccccccccceccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhh
MEM      .....

SEQ      FIFLRNLVREYLERELTMASLPFTFDVERSIDDVFMCFVFGNDFLPHLPSLEIRENAI
SEG
PRD      cchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM      .....MMMMMMMMMMMMMMMMMM.....

SEQ      DRLVNIYKNVVKHTGGYLTESGYVNLQVRQIMLAVGEVEDSI FK RK R D D E D S F R R R Q K E
SEG
PRD      hhhhhhhhhhhhhccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM      .....xxxxxxx.....

SEQ      KRKRMRKDQPAFTPSGILTPHALGSRNSPGSQVASNPQAAAYEMRMQNNSSPSISPTSF
SEG
PRD      xxxxxxxx.....xxxxxxxxxxxxxxxxxxxx
MEM      hhhhhhhhhccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
.....

SEQ      TSDGSPSPLGGIKRKAEDSDSEPEPDNVRLWEAGWKQRYRYKNKFDVDAADEKFRKRVVQ
SEG
PRD      xx.....xxxxxxxxxxxxx.....
MEM      cccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
.....

SEQ      SYVEGLCWVLRYYYGCASWKWYYPFHYPASFDFEGIADMPDSEFKGTKPKFKPLEQLMG
SEG
PRD      hhhhhhheeeeeccccccccccccccccccccccccccccccccccccccccccccccccccccchhhhh
MEM      .....

SEQ      VFPAASGNFLPSPWRKLSMDPDSIIDFYPEDFAIDLNGKKYAWQGVALLPFVDERRLRA
SEG
PRD      hccccccccccccccccccccccccceccccceccccccccceeeeeeeccccchhhhhhh
MEM      .....

SEQ      ALEEVYPDLTPEETRNRSLGGDVLVFGKHHPLHDFILELYQTGSTPEVVPPELCHGIQG
SEG
PRD      hhhhccccchhhhhhhccccceeeeeccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM      .....

SEQ      KFSLDEEAILPDQIVCSPVPMRLDLTQNTVVSINFKDPQFAEDYIFKAVMLPGARKPAAV
SEG

```

```

PRD      cccccceccccceccccccccccccccccccccccccchhhheccccccccce
MEM      .....

SEQ      LKPSDWEKSSNGRQWKPLGFNRDRRPVHLDQAAFRTLGHVMPRGSGTGIYSNAAPPVVT
SEG      .....
PRD      eccccccccccccccccccccccccccccchhhhhhhhhcccccccccccccccc
MEM      .....

SEQ      YQGNLYRPLLRGQAQIPKLMSNMRPQDSWRGPPPLFQQQRFDRGVGAEP LLPWNRMLQTQ
SEG      .....
PRD      cccccchhhhhccccchhhhhccccccccccccccccchhhhhccccccccccccchhhhh
MEM      .....

SEQ      NAAFQPNQYQMLAGPGGYPPRRDRGGRGYPREGRYPLPPPSGRYNWN
SEG      .....xxxxxxxxxxxxxxxxxxxxx.....
PRD      hcccccccccecccccccccccccccccccccccccccccccccccccccc
MEM      .....

```

Prosites for DKFZphtes3_2m18.3

PS00001	190->194	ASN_GLYCOSYLATION	PDOC00001
PS00001	247->251	ASN_GLYCOSYLATION	PDOC00001
PS00001	468->472	ASN_GLYCOSYLATION	PDOC00001
PS00001	477->481	ASN_GLYCOSYLATION	PDOC00001
PS00002	826->830	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	675->679	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	11->14	PKC_PHOSPHO_SITE	PDOC00005
PS00005	116->119	PKC_PHOSPHO_SITE	PDOC00005
PS00005	413->416	PKC_PHOSPHO_SITE	PDOC00005
PS00005	559->562	PKC_PHOSPHO_SITE	PDOC00005
PS00005	613->616	PKC_PHOSPHO_SITE	PDOC00005
PS00005	674->677	PKC_PHOSPHO_SITE	PDOC00005
PS00005	868->871	PKC_PHOSPHO_SITE	PDOC00005
PS00005	944->947	PKC_PHOSPHO_SITE	PDOC00005
PS00006	63->67	CK2_PHOSPHO_SITE	PDOC00006
PS00006	331->335	CK2_PHOSPHO_SITE	PDOC00006
PS00006	499->503	CK2_PHOSPHO_SITE	PDOC00006
PS00006	501->505	CK2_PHOSPHO_SITE	PDOC00006
PS00006	541->545	CK2_PHOSPHO_SITE	PDOC00006
PS00006	573->577	CK2_PHOSPHO_SITE	PDOC00006
PS00006	583->587	CK2_PHOSPHO_SITE	PDOC00006
PS00006	619->623	CK2_PHOSPHO_SITE	PDOC00006
PS00006	624->628	CK2_PHOSPHO_SITE	PDOC00006
PS00006	670->674	CK2_PHOSPHO_SITE	PDOC00006
PS00006	723->727	CK2_PHOSPHO_SITE	PDOC00006
PS00006	784->788	CK2_PHOSPHO_SITE	PDOC00006
PS00007	659->667	TYR_PHOSPHO_SITE	PDOC00007
PS00008	125->131	MYRISTYL	PDOC00008
PS00008	375->381	MYRISTYL	PDOC00008
PS00008	450->456	MYRISTYL	PDOC00008
PS00008	600->606	MYRISTYL	PDOC00008
PS00008	825->831	MYRISTYL	PDOC00008
PS00008	829->835	MYRISTYL	PDOC00008
PS00008	926->932	MYRISTYL	PDOC00008
PS00009	638->642	AMIDATION	PDOC00009
PS00009	934->938	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3_2m18.3)

DKFZphtes3_2m20

group: testes derived

DKFZphtes3_2m20 encodes a novel 183 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

group: unknown

DKFZphtes3_2m20 encodes a novel

amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

EST hits are only from testis or uterus librarrys
remaining intron in3' UTR see EST-BLAST

Sequenced by EMBL

Locus: unknown

Insert length: 1341 bp

Poly A stretch at pos. 1320, polyadenylation signal at pos. 1300

```
1  GCAATCCAGG AGCTGAATGG TAACTCTTCC ACAAGCGAAA ACTGTTCGTG
51  AATACAAGCA AAAGGCCCCC CAAGAGGACC CCTGATATGA TCCAGCAGCC
101 TCGGGCCCCG CTGGTGTGGG AGAAGGCTTC TGGTGAAAGGA TTTGGCAAAA
151 CCGCCCGCTAT TATACAGCTC GCTCCTAAAG CTCCTGTGTA CCTGTGTGAG
201 ACAGAGAAAC TGAGGGCAGC CTTCTTTGCA GTCCCGTTGG AAATGAGAGG
251 GTCCTTCCTG GTGCTGCTCC TGAGGGAATG CTTCCGAGAC CTGAGCTGGC
301 TGGCACTCAT CCATAGCGTC CGTGGGGAGG CGGGGCTGCT GGTGACGAGT
351 ATCGTCCCGA AGACCCCGTT TTTCTGGGCC ATGCACATCA CTGAGGCTCT
401 GCACCAGAAC ATGCAGGCTC TGTTTAGCAC CCTGGCTCAG GCGGAGGAGC
451 AGCAGCCCTA CCTGGAGGCT CCACCGTTAT GCGCGGGACT CGCTGTCTGG
501 CAGAGTACCA CCTGGGGGAT TATGGACACG CCTGGAACAG GTGTTGGGTG
551 CTGGACAGGG TGGACACCTG GGCTGTGGTC ATGTTTATG ATTTTGGACA
601 GTTGGCCACC ATCCCTGTGC AGTCTCTGCG CCAGCTAGAC AGCGACGACT
651 TCTGGACCAT CCCACCCCTG ACTCAGCCAT TCATGCTGGA GAAAGACATT
701 TTAGATTTCG ATGAGGTTGT CCATCGAATC CTCAAAGGGA AAATCACTGG
751 TGCCTTTGAA TCGGCGGTAA CTGCTCCTGC ATCTAACTTG GCTGTTGTCC
801 CTCCACTCCT GCCCTTGGGG TGTCTGCAGC AGGCTGCTGC CTAGGCCTGG
851 ACACATTGCA CATCCTAAAG TTTGAAGAGT CTAATAACG GGGCTTCCCT
901 CAGCATGTTC CCTCTCCTGT TTGCCACGGA TCCAGAGCCA CCTGCCCTGT
951 CTTCTCGTAC CCCTTTCACT CTTGAGGCCT GGGAGGTGAA AAAGGCCAGA
1001 CTGTGCCCAG GATTGATTCA ATTTTGCTTT TACTCCCAGC TTCCCTCTCA
1051 AAAGAGAGTG AAGTCTCATT TGTCATGTGT CTTCAAGTCC CCAACTTGGC
1101 ATGAACATTT GAACCAACA TAGGAAACTA CCATTAGGTT GAAAGCCTGA
1151 GGCAGCTGGG ATGGTCTTTC TTGTGTCTCT TCTTTGACC CCAGAGCATG
1201 ATATAAGTGG TCCTAACAGA TTCTGGATAA TGGAGAAGCC CTCTGCTGGT
1251 TTTCTGGCA TTCCATGTAG AATAGGTAGA GAATATTTAA CCAATGAGCA
1301 AATAAATGTT GGCATGTTTC ATGAAAAAAA AAAAAAAAAA A
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 479 bp to 841 bp; peptide length: 121
 Category: questionable ORF
 Classification: no clue

1 MRGTRCLAEY HLG DYGHAWN RCWVLD RVD T WAVVMFIDFG QLATIPVQSL
 51 RQLDSDDFWT IPPLTQPFML EKDILSSYEV VHRILKGKIT GALNSAVTAP
 101 ASNLAVVPPL LPLGCLQQAA A

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2m20, frame 2

No Alert BLASTP hits found

Peptide information for frame 3

ORF from 87 bp to 635 bp; peptide length: 183
 Category: putative protein
 Classification: no clue

1 MIQQPRAPLV LEKASGEGFG KTAIIQLAP KAPVDLCETE KLRAAFFAVP
 51 LEMRGSFLVL LLRECFRDLs WLALHSVRG EAGLLVTSIV PKTPFFWAMH
 101 ITEALHQNMQ ALFSTLAQAE EQQYLEAPP LCAGLAVWQS TTWGIMDTPG
 151 TGVGCWTGWT PGLWSCSLIL DSWPPSLCSL CAS

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2m20, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_2m20, frame 2

Report for DKFZphtes3_2m20.2

[LENGTH] 121
 [MW] 13436.69
 [pI] 5.81
 [KW] Alpha_Beta

SEQ MRGTRCLAEYHLGDYGHAWNRCWVLD RVD T WAVVMFIDFGQLATIPVQSLRQLDSDDFWT
 PRD ccchhhhhcc

SEQ IPPLTQPFMLEKDILSSYEVVHRILKGKITGALNSAVTAPASNLA VVPPLPLGCLQQAA
 PRD cccccchhhhhhhcchhhhhhhhhccccchhhhhcccccccccccccccccccccccccccc

SEQ A
 PRD c

(No Prosite data available for DKFZphtes3_2m20.2)

(No Pfam data available for DKFZphtes3_2m20.2)

Pedant information for DKFZphtes3_2m20, frame 3

Report for DKFZphtes3_2m20.3

[LENGTH] 183
 [MW] 19971.49
 [pI] 5.31
 [KW] Alpha_Beta

SEQ MIQQPRAPLVLEKASGEGFGKTAIIQLAPKAPVDLCETEKLRAAFFAVPLEMRGSFLVL
PRD cccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhcchhhhhh

SEQ LLRECFRDLSWLALIHSVRGEAGLLVTSIVPKTPFFWAMHITEALHQNMQUALFSTLAQAE
PRD hhhhhhcchhhhhhhhhhhccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ EQQPYLEAPPLCAGLAVWQSTTWGIMDTPGTGVGCWTGWTPGLWSCSLILDSWPPSLCSL
PRD hhhcc

SEQ CAS
PRD ccc

(No Prosite data available for DKFZphtes3_2m20.3)

(No Pfam data available for DKFZphtes3_2m20.3)

DKFZphtes3_2n9

group: testes derived

DKFZphtes3_2n9 encodes a novel 184 amino acid protein with very weak similarity to Homo sapiens PAC clone DJ0771P04 from 7q11.21-q11.23.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

on genomic level encoded by HS1186N24, no splice pattern but EST matches

Sequenced by EMBL

Locus: unknown

Insert length: 1000 bp

Poly A stretch at pos. 988, polyadenylation signal at pos. 970

```
1 CAACCTTTTAA AAGATGTGAA TTGGACAGCC AGACTTGCTT ATTTGTCTGA
51 TATCTTCAGT ATTTTAAAT GATCTTAATG CTCTATGCA AGGGAAGAAT
101 GCAACTTATT TTCAATGGC AGATAAAGTT GAAGGACAAA AACAGAAGTT
151 AGAAGCTTGG AAAACAGAA TTCTACAGA TTGTTATGAC ATGTTTCATA
201 ATTTAACAAC AATTATCAAT GAAGTAGGTA ATGATCTTGA TATTGCACAT
251 CTGCGAAAAG TTATCAGTGA ACATCTTACA AATTTGTTAG AATGTTTGA
301 ATTTTATTTT CCATCAAAAG AAGATCCACG CATAGGAAAT TTGTGGATCC
351 AAAATCCATT TCTTTCATCA AAAGATAACT TAAATTTAAC TGTAACCTTA
401 CAGGATAAGT TGTGGAAGCT GGCTACCGAC GAAGGATTGA AAATCAGTTT
451 TGAAAATACA GCATCACTTC CTTCAATTTG GATAAAAGCT AAAAATGACT
501 ATCCTGAGCT TGCTGAGATT GCTTTAAAT TGCTGCTTCT TTTCCCTCA
551 ACATACCTCT GTGAGACCGG ATTCTCTACT TTAAGTGTTA TTTAAACAAA
601 ACATAGAAAC AGTTTAAATA TACATTATCC CCTGAGGTAG CATGTGCATC
651 AATCCAACCT AGATTAGACA AATTAACAAG CAAGAAGCAA GCTCACTTAT
701 CACATTAATA GCTTTAAATA TTGATATGTA AGGTATTGGT TCAAAGTATG
751 CATATAAGCA TTGAGTGTGA GGAATTGCT ATTTCACTTT AAACCTTCTG
801 TCTAGTTACA GTTATGGAAG TATGAGAAGT TATGAGTGAA ACAGCAATTT
851 TCTATATAAA TTGCCTATAT GTATATTTTC AATTAAGAAT GTGTACAGTT
901 TTTATAATTC TATTTTTCCT CATATTGTC GTATTTATTA AAATATAATT
951 TTAATCTGTG TGATTCTAAT ATTAACACAT TTGATCTTAA AAAAAAAAAA
```

BLAST Results

Entry HS1186N24 from database EMBLNEW:
Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 1186N24
Score = 4921, P = 5.8e-215, identities = 989/992

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 86 bp to 637 bp; peptide length: 184
Category: similarity to unknown protein
Classification: no clue

```
1 MQGKNATYFS MADKVEGQKQ KLEAWKNRIS TDCYDMFHNL TTIINEVGND
51 LDIAHLRKVI SEHLTNLLEC FEFYFPSKED PRIGNLWIQN PFLSSKDNLN
101 LTVTLQDKLL KLATDEGLKI SFENTASLPS FWIKAKNDYP ELAEIALKLL
151 LLFPSTYLCE TGFSTLSVIK TKHRNSLNIH YPLR
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2n9, frame 2

TREMBLNEW:AC004883_3 gene: "WUGSC:H_DJ0771P04.2"; Homo sapiens PAC clone DJ0771P04 from 7q11.21-q11.23, complete sequence., N = 1, Score = 94, P = 0.042

>TREMBLNEW:AC004883_3 gene: "WUGSC:H_DJ0771P04.2"; Homo sapiens PAC clone DJ0771P04 from 7q11.21-q11.23, complete sequence.
Length = 533

HSPs:

Score = 94 (14.1 bits), Expect = 4.3e-02, P = 4.2e-02
Identities = 39/177 (22%), Positives = 75/177 (42%)

Query: 1 MQGKNATYFSMADKVEGQKQKLEAWKNRISTDCYDMFHNLTIIINEVGNDLDIAHLRKV 59
+QG + M D + KL W+ ++ + F L + L+ I + ++
Sbjct: 354 LQGHSQIVTQMYDLIRAFIAKLCLWETHLTRNNLAHFPTLKLASRNESDGLNYIPKIAEL 413

Query: 60 ISEHLTNLLECFEFYFPSKEDPRIGNLWIQNPFLSSKDNLNLTVTLQDKLLKLATDEGLK 119
+E L + F+ Y + + + +PF + D+++ LQ +++ L + LK
Sbjct: 414 KTEFQKRLSD-FKLY---ESELTL----FSSPFSTKIDSVH--EELQMEVIDLQCNTVLK 463

Query: 120 ISFENTASLPFSFWIKAKNDYPXXXXXXXFPSTYLCETGFSTLSVIKTKHRNSL 177
++ +P F+ YP F STY+CE FS + + KTK+ + L
Sbjct: 464 TKYDKVG-IPFYKYLWGSYPKYKHHCAKILSMFGSTYICEQLFSIMKLSKTKYCSQL 520

Pedant information for DKFZphtes3_2n9, frame 2

Report for DKFZphtes3_2n9.2

{LENGTH}	184
{MW}	21203.53
{pI}	6.52
{KW}	Alpha Beta
{KW}	LOW_COMPLEXITY 6.52 %

SEQ	MQGKNATYFSMADKVEGQKQKLEAWKNRISTDCYDMFHNLTIIINEVGNDLDIAHLRKVI
SEG
PRD	ccccccchhhhhhhhhhhhhhhhhhhhhcchhhhhccccccccccccchhhhhhhhh
SEQ	SEHLTNLLECFEFYFPSKEDPRIGNLWIQNPFLSSKDNLNLTVTLQDKLLKLATDEGLKI
SEG
PRD	hhhhhhhhhhhhcc
SEQ	SFENTASLPFSFWIKAKNDYPELAELKLLLPSTYLCETGFSTLSVIKTKHRNSLNIH
SEGXXXXXXXXXXXX.....
PRD	eeccccccccccccccccchhhhhhhhhhhhhcccccccccccccccccccccccc
SEQ	YPLR
SEG
PRD	cccc

(No Prosite data available for DKFZphtes3_2n9.2)

(No Pfam data available for DKFZphtes3_2n9.2)

DKFZphtes3_30f4

group: testes derived

DKFZphtes3_30f4 encodes a novel 192 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

Sequenced by LMU

Locus: /map="717.2-8 cR from top of Chr8 linkage group"

Insert length: 1388 bp

Poly A stretch at pos. 1330, polyadenylation signal at pos. 1310

```

1  CACTGAGCCC TCCTCAGATG GTTAGTGGCT TCCAACAGCC ATCAGGAGTG
51 TTTCTTGAAT GCCCCAGGTG TGGAGGACTT GGTCTGTGAC CACCTAGAAC
101 CCCAGAGCTG AACAGGAAGC CGTCCCTGCA GCAACAAGAG GGCTGGAAGG
151 GGGAGCTGCA GGCCACCCTC GGCTCTCCCA CTGCTGGGGC GGTGATGTTT
201 GGGTGACATG TTTGAAAAAT ACTCTTAAAG ATACCAACTG TTCCTTATA
251 TGGCTAATGG TTTGTGCAGC CACCAGCGAT GCGGGCCCTT ATTAGAGACC
301 AGGTTTGTTA AAACACCAA TATTGCTGTC CACACTAGAC ATTAACCGGC
351 TTCAGAAAAG ATGGACACCT TTTCCACGCG TGTTCGCTT CTTAACTTTG
401 GTCCAGCTTT AGCCACCACA CAGCGTGTGA GGGACTGCTG CTGCGGAGTC
451 AGCCTCGTTT GTCCCTCCGC CTCCACCCAG CATGCGCCGC TTCTGAGAGA
501 CACCAGCTCC CTGCCTCCAA GCCTGGTGCC ACAGGCCTGT CGTGAGGGAC
551 CCCTGCTTCC GAGAGCTCCT GGGGGGGTTC TGCCCTTCAC CACCTGGGAG
601 AGGTGTCACT TCAGTTCCGA GTTGAACAAG GCCCGTGCAC ACAGCATGTT
651 GGGGGGCCAG CCCAAAGTTC TTGTACCTC CTCATGCAAA GCCAGCCATC
701 ACCCTCCGGC CAGAGCTCAA GGTGGCCCTT TGGCCAGCCC CTCCTTGGGT
751 CCTCCAGGAG GACTGAGCAC CCCTCCTAGC GGCATCCCTT GCCCTCCACA
801 GTGCTGCCAG GGGCACGTCG CTCTGTGCCG TGGACTGAGA CCATCCCCTG
851 GTGACAGAAT GACCCGTTTG TTGGAAATGC CTCGTTGCCA GAGAACTCC
901 CCAGGCATCT CGGAACGAAA CTATTTAGTT CCATTGTGAA CTGGCCACGG
951 GACAGCTTTT TATCAACTTA TTAAGTTGGA GCACGTGAAT CGCGCTTGCT
1001 GAGTTAGCAG TGGTGGTAAG CGTGTGTTAA ACACATAATG TTACGTTTTA
1051 GGAGAGAGAG GTCGTAAGGA AGTGTCGTGT CGCTCATGAC TCTCTTCTAT
1101 TAGTTGGGTA ACAGTGGCCT CATGTTTGTG TCTGTGTGTA CACAGAGCCC
1151 TTAGGTTCTG CTCTGTTTCT TTGCCAGGTG AATGTTTGTG GCATGCGCTG
1201 CTGTCCGCGC CCCTCTGTCC TCGCCAGGGT TCAGCTGTGC GCGGCCCTGA
1251 TTTCTCCAT GCACACAGAA CCTCCTTGTG TCTGTTTCTC TGTTCTCTGT
1301 TGGCTGACTC AATAAACTTT TCCCTCTGAC ATGAAAAAAA AAAAAAAGAA
1351 AAAAAAAGAA AAAAAAAGAA AAAAAAAGAA AAAAAAAGAA

```

BLAST Results

Entry HS548358 from database EMBL:

human STS EST67250.

Score = 2126, P = 1.5e-89, identities = 444/472

Entry HS670351 from database EMBL:

human STS WI-18501.

Score = 2089, P = 7.1e-88, identities = 445/476

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 361 bp to 936 bp; peptide length: 192

Category: putative protein

Classification: no clue

1 MDTFSHAVSL LNFGPALATT QVRDCCCGV SLVCPSSHQ HAPLLRDTSS
51 LPPSLVPQAC REGPLLPRAP GGVLPTTWE RCQFSSELNK ARAHSMGLAQ
101 PKVLVTSSCK ASHHPPARAQ GGPLASPSLG PPGGLSTPPS GIPCPPQCCQ
151 GHVALCRGLR PSPGDRMTRL LEMPRCQRNS PGISERNYLVL PL

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_30f4, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_30f4, frame 1

Report for DKFZphtes3_30f4.1

[LENGTH] 192
[MW] 20281.56
[pI] 9.21
[BLOCKS] BL01013C Oxysterol-binding protein family proteins
[KW] All Alpha
[KW] LOW_COMPLEXITY 10.94 %

SEQ MDTFSHAVSLLNFGPALATTQVRDCCCGVSLVCPSSHQHAPLLRDTSSLPPSLVPQAC
SEG
PRD cccchhhheeeccccchhhhhhccccceeecccccccccccccccccccccccccc

SEQ REGPLLPRAPGGVLPTTWERCQFSSELNKARAHSMGLAQPKVLVTSSCKASHPPARAQ
SEG
PRD cccccccccccccccccccchhhhhhhhhhhhhccccceeecccccccccccccc

SEQ GGPLASPSLGPPGGLSTPPSGIPCPPQCCQGHVALCRGLRPSPGDRMTRLLEMPCQRNS
SEG xx
PRD cccccccccccccccccccccccccccccccchhhhhhhhhccccccchhhhhcccccc

SEQ PGISERNYLVLPL
SEG
PRD cccccccccccc

(No Prosite data available for DKFZphtes3_30f4.1)

(No Pfam data available for DKFZphtes3_30f4.1)

DKFZphtes3_35b4

group: cell cycle

DKFZphtes3_35b4 encodes a novel 1780 amino acid protein which is C-terminal identical to human M-phase phosphoprotein-1 (MPP1).

The novel protein contains a N-terminal Pfam kinesin motor domain and a ATP/GTP-binding site motif A (P-loop). MPP1 is expressed and phosphorylated in the metaphase. Therefore the novel protein seems to be involved in the mitotic spindle during cell division.

The new protein can find application in modulation of the mitotic spindle.

"M-phase phosphoprotein-1" extension

motor protein

Sequenced by DKFZ

Locus: /map="750_H_1; 758_H_7; 759_C_9; 847_D_4; 906_D_1; 931_D_3; 944_C_1; 750_G_12; 800_A_11; 512.1 cR from top of Chr10 linkage group"

Insert length: 6284 bp

No poly A stretch found, no polyadenylation signal found

```

1 ATCGCAGTGC TGCTCGCGGG TCTGGCTAGT CAGGCGAAGT TTGCAGAATG
51 GAATCTAATT TTAATCAAGA GGGAGTACCT CGACCATCTT ATGTTTTTAG
101 TGCTGACCCA ATTGCAAGGC CTTGAGAAAT AAATTTTCGAT GGCATTAAAGC
151 TTGATCTGTC TCATGAATTT TCCTTAGTTG CTCCAATAC TGAGGCAAAAC
201 AGTTTCGAAT CTAAAGATTA TCTCCAGGTT TGTCTTCGAA TAAGACCATT
251 TACACAGTCA GAAAAAGAAC TTGAGTCTGA GGGCTGTGTG CATATTCTGG
301 ATTCACAGAC TGTGTGCTG AAAGAGCCTC AATGCATCCT TGGTCGGTTA
351 AGTGAAAAAA GCTCAGGGCA GATGGCACAG AAATTCAGTT TTTCCAAGGT
401 TTTTGGCCCA GCAACTACAC AGAAGGAATT CTTTCAGGGT TGCATTATGC
451 AACCAGTAAA AGACCTCTTG AAAGGACAGA GTCGTCTGAT TTTTACTTAC
501 GGGCTAACCA ATTCAGGAAA AACATATACA TTTCAAGGGA CAGAAGAAAA
551 TATTGGCATT CTGCCTCGAA CTTTGAATGT ATTATTGAT AGTCTTCAAG
601 AAAGACTGTA TACAAAGATG AACCTTAAAC CACATAGATC CAGAGAATAC
651 TTAAGGTTAT CATCAGAACA AGAGAAAGAA GAAATTGCTA GCAAAAGTGC
701 ATTGCTTCGG CAAATTAAAG AGGTTACTGT GCATAATGAT AGTGATGATA
751 CTCTTTATGG AAGTTTAACT AACTCTTTGA ATATCTCAGA GTTTGAAGAA
801 TCCATAAAAG ATTATGAACA AGCCAACCTG AATATGGCTA ATAGTATAAA
851 ATTTTCTGTG TGGGTTTCTT TCTTTGAAAT TTACAATGAA TATATTATATG
901 ACTTATTTGT TCTGTATCA TCTAAATTCC AAAAGAGAAA GATGCTGCGC
951 CTTTCCCAAG ACGTAAAGGG CTATTCCTTT ATAAAAGATC TACAATGGAT
1001 TCAAGTATCT GATTCCAAAG AAGCCTATAG ACTTTTAAAA CTAGGAATAA
1051 AGCACCAGAG TGTTGCCTTC ACAAATTTGA ATAATGCTTC CAGTAGAAGT
1101 CACAGCATAT TCACTGTTAA AATATTACAG ATTGAAGATT CTGAAATGTC
1151 TCGTGTAATT CGAGTCAGTG AATTATCTTT ATGTGATCTT GCTGGTTCAG
1201 AACGAACTAT GAAGACACAG AATGAAGGTG AAAGGTTAAG AGAGACTGGG
1251 AATATCAACA CTTCTTTATT GACTCTGGGA AAGTGATTA ACGTCTTGAA
1301 GAATAGTGAA AAGTCAAAGT TTCAACAGCA TGTGCCTTTC CGGAAAGTA
1351 AACTGACTCA CTAATTTCAA AGTTTTTTTA ATGGTAAAGG GAAAAATTGT
1401 ATGATTGTCA ATATCAGCCA ATGTTATTTA GCCTATGATG AAACACTCAA
1451 TGTATTGAAG TTCTCCGCCA TTGCACAAA AGTTTGTGTC CCAGACACTT
1501 TAAATTCCTC TCAAGATAAA TTATTTGGAC CTGTCAAATC TTCTCAAGAT
1551 GTATCACTAG ACAGTAATTC AAACAGTAAA ATATTAATG TAAAAAGAGC
1601 CACCATTTC A TGGGAAAATA GTCTAGAAGA TTTGATGGAA GACGAGGATT
1651 TGGTTGAGGA GCTAGAAAAC GCTGAAGAAA CTCAAAATGT GGAACATAAA
1701 CTTCTTGATG AAGATCTAGA TAAAACATTA GAGGAAAATA AGGCCTTCAT
1751 TAGCCACGAG GAGAAAAGAA AACTGTTGGA CTTAATAGAA GACTTGAAAA
1801 AAAAACTGAT AAATGAAAAA AAGGAAAAAT TAACCTGGA ATTTAAAT
1851 CGAGAAGAAG TTACACAGGA GTTTACTCAG TATTGGGCTC AACGGGAAGC
1901 TGACTTTAAG GAGACTCTGC TTCAAGAACG AGAGATATTA GAAGAAAATG
1951 CTGAACGTCG TTTGGCTATC TTCAAGGATT TGGTTGGTAA ATGTGACACT
2001 CGAGAAGAAG CAGCGAAAGA CATTTGTGCC ACAAAGTTG AAAGTGAAGA
2051 AGCTACTGCT TGTTTAGAAC TAAAGTTTAA TCAAAATTA GCTGAATTAG
2101 CTAAACCCTA AGGAGAATTA ATCAAAACCA AAGAAGAGTT AAAAAAGAGA
2151 GAAATGAAT CAGATTCATT GATTCAAGAG CTTGAGACAT CTAATAAGAA
2201 AATAATTACA CAGATCAAAA GAATTTAAAG ATTGATAAAT ATAATTGATC
2251 AAAAAAGAAG TACTATCAAC GAATTTTACA ACCTAAAGTC TCATATGGAA
2301 AACACATTTA AATGCAATGA CAAGGCTGAT ACATCTTCTT TAATAATAAA
2351 CAATAAATTG ATTTGTAATG AAACAGTTGA AGTACCTAAG GACAGCAAAAT
2401 CTAAATCTCT TTCAGAAAAG AAAAGAGTAA ATGAAAATGA ACTTCAGCAA
2451 GATGAACCA CAGCAAAGAA AGGGTCTATC CATGTTAGTT CAGCTATCAC
2501 TGAAGACCAA AAGAAAAGTG AAGAAGTGGC ACCGAACATT CGAGAAATTG
2551 AAGACATCAG AGTTTACAAA GAAAATAATG AAGGACTGAG AGCATTTTTA

```

```

2601 CTCACTATTG AGAATGAAC TAAAAATGAA AAGGAAGAAA AAGCAGAATT
2651 AAATAAACAG ATTGTTCAAT TTCAGCAGGA ACTTTCTCTT TCTGAAAAAA
2701 AGAATTAAAC TTAAAGTAAA GAGGTCCAAC AAATTCAGTC AAATTATGAT
2751 ATTGCAATTG CTGAATTACA TGTGCAGAAA AGTAAAAATC AAGAACAGGA
2801 GGAAGAGATC ATGAAATTGT CAAATGAGAT AGAAACTGCT ACAAGAAGCA
2851 TTACAAATAA TGTTTCACAA ATAAAAATTAA TGCACACGAA AATAGACGAA
2901 CTACGTACTC TTGATTCACT TTCTCAGATT TCAACATAG ATTGCTCAA
2951 TCTCAGGGAT CTGTCAAATG GTTCTGAGGA GGATAATTG CCAAATACAC
3001 AGTTAGACCT TTTAGGTAAT GATTATTTGG TAAGTAAGCA AGTTAAAGAA
3051 TATCGAATTC AAGAACCCTA TAGGGAAAAAT TCTTTCCACT CTAGTATTGA
3101 AGCTATTTGG GAAGAATGTA AAGAGATTGT GAAGGCCTCT TCCAAAAAAA
3151 GTCATCAGAT TGAGGAACTG GAACAACAAA TTGAAAAATT GCAGGCAGAA
3201 GTTAAAGGCT ATAAGGATGA AAACAATAGA CTAAGGAGA AGGAGCATAA
3251 AAACCAAGAT GACCTACTAA AAGAAAAAGA AACTCTTATA CAGCAGCTGA
3301 AAGAAGAATT GCAAGAAAAA AATGTTACTC TTGATGTTCA AATACAGCAT
3351 GTAGTTGAAG GAAAGAGAGC GCTTTCAGAA CTTACACAAG GTGTTACTTG
3401 CTATAAGGCA AAAATAAAGG AACTTGAAAC AATTTTAGAG ACTCAGAAAG
3451 TTGAACGTAG TCATTCAAGC AAGTTAGAAC AAGACATTTT GGAAAGGAA
3501 TCTATCATCT TAAAGCTAGA AAGAAATTTG AAGGAATTTT AAGAACATCT
3551 TCAGGATTCT GTCAAAAAACA CCAAGATTTT AAATGTAAAG GAACTCAAGC
3601 TGAAAGAAGA AATCACACAG TTAACAAATA ATTTGCAAGA TATGAACAT
3651 TTACTTCAAT TAAAAGAAGA AGAAGAAGAA ACCAACAGGC AAGAAACAGA
3701 AAAATTGAAA GAGGAACTCT CTGCAAGCTC TGCTCGTACC CAGAATCTGA
3751 AAGCAGATCT TCAGAGGAAG GAAGAAGATT ATGCTGACCT GAAAGAGAAA
3801 CTGACTGATG CCAAAAAGCA GATTAAGCAA GTACAGAAAG AGGTATCTGT
3851 AATGCGTGAT GAGGATAAAT TACTGAGGAT TAAATTAAT GAACTGGAGA
3901 AAAAGAAAAA CCAGTGTTCT CAGGAATTAG ATATGAAGCA CGGAACCAT
3951 CAGCAACTCA AGGAGCAGTT AAATAATCAG AAAGTGGAA GAGCTATACA
4001 ACAGTATGAG AGAGCATGCA AAGATCTAAA TGTTAAAGAG AAAATAATTG
4051 AAGACATGCG AATGACACTA GAAGAACAGG AACAACTCA GGTGAACAG
4101 GATCAAGTGC TTGAGGCTAA ATTAGAGGAA GTTGAAGGC TGCCACAGA
4151 ATTGGA AAAA TGGAAGGAAA AATGCAATGA TTTGGAACC AAAACAATC
4201 AAAGGTCAAA TAAAGAACAT GAGAACAAACA CAGATGTGCT TGGAAAGCTC
4251 ACTAATCTTC AAGATGAGTT ACAGGAGTCT GAACAGAAAT ATAATGCTGA
4301 TACAAAGAAA TGGTTAGAAG AAAAAATGAT GCTTATCACT CAAGCGAAA
4351 AAGCAGAGAA TATACGAAAT AAAGAGATGA AAAAATATGC TGAGGACAGG
4401 GAGCGTTTTT TTAAGCAACA GAATGAAATG GAAATACTGA CAGCCCAGCT
4451 GACAGAGAAA GATAGTGACC TTCAAAGTG GCGAGAAGAA CGAGATCAAC
4501 TGGTTGCAGC TTTAGAAATA CAGCTAAAAG CACTGATATC CAGTAATGTA
4551 CAGAAAGATA ATGAAATTGA ACAACTAAAA AGGATCATAT CAGAGACTTC
4601 TAAAAATGAA ACACAAATCA TGGATATCAA GCCCAACGT ATTAGTTGAG
4651 CAGATCTGTA CAACTTCAA ACTGAACCTC TATCGACAAG TTTGAAATT
4701 TCCAGAAATA AAATAGAGGA TGGATCTGTA GTCCTTGACT CTGTGTAAGT
4751 GTCAACAGAA AATGATCAAA GCACTCGATT TCCAAACCT GAGTTAGAGA
4801 TTCAATTTAC ACCTTTACAG CCAACAAAAA TGGCAGTGAA ACACCCTGGT
4851 TGTACCACAC CAGTGACAGT TGAGATTCCC AAGGCTCGGA AGAGGAAGAG
4901 TAATGAAATG GAGGAGGACT TGGTGAAATG TGAAATAAG AAGAATGCTA
4951 CACCCAGAAC TAATTTGAAA TTTCTTATTT CAGATGATAG AAATCTTCTT
5001 GTCAAAAAGG AACAAAAGGT TGCCATACGT CCATCATCTA AGAAAACATA
5051 TTCTTTACGG AGTCAGGCAT CCATAATTGG TGTAAACCTG GCCACTAAGA
5101 AAAAAAGAGG AACACTACAG AAATTTGGAG ACTTCTTACA ACATTCTCCC
5151 TCAATTTCTT AATCAAAAAGC AAAGAAGATA ATTGAAACAA TGAGCTCTTC
5201 AAAGCTCTCA AATGTAGAAG CAAGTAAAGA AAATGTGCTT CAACCAAAAC
5251 GAGCCAAACG GAAATTATAC ACAAGTGAAA TTTCACTCTC TATTGTATATA
5301 TCAGGCCAAG TGATTTTAAT GGACCAAGAA ATGAAGGAGA GTGATCACCA
5351 GATTATCAAA CGACGACTTC GAACAAAAAC AGCCAAATAA ATCACTTATG
5401 GAAATGTTTA ATATAAATTT TATAGTCATA GTCATTGGAA CTGTCATCCT
5451 GTATTGTAAA TATAAATGTA TATATTATGC ATTAAATCAC TCTGCATATA
5501 GATTGCTGTT TTATACATAG TATAATTTTA ATTCAATAAA TGAGTCAAAA
5551 TTTGTATATT TTTATAAGGC TTTTATATAA TAGCTTCTTT CAAACTGTAT
5601 TTTCCCTATTA TCTCAGACAT TGGATCAGTG AAGATCCTAG GAAAGAGGCT
5651 GTTATTCTCA TTTATTTTGC TATACAGGAT GTAATAGGTC AGGTATTTGG
5701 TTTACTTATA TTTAACAATG TCTTATGAAT TTTTCTTACT TTATCTGTTA
5751 TACAACCTGAT TTTACATATC TGTTTGGATT ATAGCTAGGA TTTGGAGAA
5801 AAGTGTTGAC AGATCACAAA ACATGTATAT ACATTATTTA GAAAAGATCT
5851 CAAGTCTTTA ATTAGAATGT CTCACCTATT TTGTAAACAT TTTGTGGGTA
5901 CATAGTACAT GTATATATTT ACGGGGTATG TGAGATGTTT TGACACAGGC
5951 ATGCAATGTG AAATACGTGT ATCATGGAGA ATGAGGTATC CATCCCCTCA
6001 AGCATTTTTC CTTTGAATTA CAGATAATCC AATTACATTC TTTAGATCAT
6051 TAAAAAATAT ACAAGTAAAG TATTATTGAT TATAGTCACT CTATTGTGCT
6101 ATCAGATAGT AGATCATTTCT TTTTATCTTA TTTGTTTGTG TACCCATTAA
6151 CCATCCCCAC CTCCCCCTGC AACCGTCAGT ACCCTTACCA GCCACTGGTA
6201 ACCATTCTTC TACTCTGTAT GCCCATGAGG TCAATTGATT TTATTTTATG
6251 ATCCCATAAA TAAATGAGAA CATGCAAAAA AAAA

```

BLAST Results

Entry HS898149 from database EMBL:
human STS WI-9217.

Score = 4247, P = 1.5e-187, identities = 855/862

Medline entries

94119956:

Cloning of cDNAs for M-phase phosphoproteins recognized by the MPM2 monoclonal antibody and determination of the phosphorylated epitope.

98101856:

Interaction of a Golgi-associated kinesin-like protein with Rab6.

95122643:

Identification and partial characterization of mitotic centromere-associated kinesin, a kinesin-related protein that associates with centromeres during mitosis.

Peptide information for frame 3

ORF from 48 bp to 5387 bp; peptide length: 1780

Category: known protein

Classification: Cell structure/motility

Prosite motifs: ATP_GTP_A (152-160)

```
1 MESNFMQEGV PRPSYVFSAD PIARPSEINF DGIKLDSHE FSLVAPNTEA
51 NSFESKDYLO VCLRIRPFTQ SEKELESEGC VHILDSQTVV LKEPQCILGR
101 LSEKSSGQMA QKFSFSKVFG PATTQKEFFQ GCIMQPVKDL LKGQSRILFT
151 YGLTNSGKTY TFQGTENIG ILPRTLNVLV DSLQERLYTK MNLKPHRSRE
201 YLRLSSEQEK EETASKSALL RQIKEVTVHN DSDDTLYGSL TNSLNISEFE
251 ESIKDYEQAN LNMANSIKFS VWVSFFEIYN EYIYDLFVPV SSKFQKRKML
301 RLSQDVKGYS FIKDLQWIVQ SDSKEAYRL LKGIKHQSVA FTKLNNASSR
351 SHSIFTVKIL QIEDSEMSRV IRVSELSLDC LAGSERTMKT QNEGERLRET
401 GNINTSLLTL GKCINVLKNS EKSKEFQHV FRESKLTHYF QSFNNGKGI
451 CMIVNISQCY LAYDETLNVL KFSIAQKVC VPDTLNSSQD KLFQPVKSSQ
501 DVSLDSNSNS KILNVKRATI SWENSLEDLM EDEDLVEELE NAEETQNVET
551 KLLDELDLKT LEENKAFISH EEKRKLDDLI EDLKKKLINE KKEKLTLEFK
601 IREEVTQFT QYWAQREADF KETLLQEREI LEENAERRLA IFKDLVGKCD
651 TREEAAKDIC ATKVETEEAT ACLELKFNQI KAEAKTKGE LIKTKEELKK
701 RENESDSLQ ELETNNKKII TQNRKIKELI NIIDQKEDTI NEFQNLKSHM
751 ENTFFKNDKA DTSSLIINN LICNETVEVP KDSKSKICSE RKRNVNELQ
801 QDEPPAKKGS IHVSSAITE QKKSEEVPRN IAEIEDIRVL QENNEGLRAF
851 LLTIENELKN EKEEKAELNK QIVHFQQLS LSEKKNLTL KEVQIQSNY
901 DIAIAELHW KSKNQEQEEK IMKLSNEIET ATRSITNNVS QIKLMHTKID
951 ELRTLDSVSQ ISNIDLLNLR DLSNGSEEDN LPNTQLDLLG NDYLVSKQVK
1001 EYRIQEPNRE NSFHSIEAI WEECKEIVKA SSKKSHQIEE LEQQIEKLQA
1051 EVKGYKDENN RLKEKEHKNQ DDLLKEKETL IQQLKEELQE KNVTLVDVQIQ
1101 HVVEGKRALS ELTQGVTCYK AKIKELETIL ETQKVERSIS AKLEQDILEK
1151 ESIIILKERN LKEFQEHLDQ SVKNTKDLNV KELKLKEEIT QLTNNLQDMK
1201 HLLQLKEEEE ETNRQETEK KEELSASSAR TQNLKADLQ KEEDYADLKE
1251 KLTDAKKQIK QVQKEVSVMR DEDKLLRIKI NELEKKKNQC SQELDMKQRT
1301 IQQLKEQLNN QKVEEAIQY ERACKDLNV EKIIEDMRMT LEEQEQTQVE
1351 QDQVLEAKLE EVERLATELE KWKEKCNLE TKNNQRSNKE HENNTDVLGK
1401 LTNLQDELQE SEQKYNADRK KWLEEKMLI TQAKEAENIR NKEKKYAED
1451 RERFFKQNE MEILTAQLTE KSDSLQKWRE ERDQLVALE IQLKALISSN
1501 VQKDNEIEQL KRIISETSKI ETQIMDIKPK RISSADPKL QTEPLSTSFE
1551 ISRNKIEDGS VVLDSECVST ENDQSTRFPK PELEIQFTPL QPNKMAVKHP
1601 GCTTPVTVEI PKARKRKSNE MEEDLVKCN KKNATPRTNL KFPISDDRNS
1651 SVKKEQKVAI RPSSKKTYSL RSQASIIGVN LATKKKEGTL QKFGDFLQHS
1701 PSILQSKAKK IETMSSSKL SNVEASKENV SQPKRAKRKL YTSEISSPID
1751 ISGQVILMDQ KMKESDHQII KRRRLRTKTA
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZptes3_35b4, frame 3

TREMBL:U93121_1 product: "M-phase phosphoprotein-1"; Human M-phase phosphoprotein-1 mRNA, partial cds., N = 1, Score = 3743, P = 0

PIR:A36881 MPM2-reactive phosphoprotein 1 - human (fragment), N = 2,
Score = 2808, P = 2.5e-294

TREMBL:AF070672_1 product: "rabkinesin6"; Homo sapiens rabkinesin6
mRNA, complete cds., N = 2, Score = 680, P = 2.6e-99

>TREMBL:U93121_1 product: "M-phase phosphoprotein-1"; Human M-phase
phosphoprotein-1 mRNA, partial cds.
Length = 753

HSPs:

Score = 3743 (561.6 bits), Expect = 0.0e+00, P = 0.0e+00
Identities = 752/753 (99%), Positives = 753/753 (100%)

Query: 1028 VKASSKKSHQIEELEQQIEKLQAEVKGKDYDENNRLEKEHEKNQDDLLKEKETLIQQLKEE 1087
VKASSKKSHQIEELEQQIEKLQAEVKGKDYDENNRLEKEHEKNQDDLLKEKETLIQQLKEE
Sbjct: 1 VKASSKKSHQIEELEQQIEKLQAEVKGKDYDENNRLEKEHEKNQDDLLKEKETLIQQLKEE 60

Query: 1088 LQEKNVTLDVQIQHVVEGKRALSELTVGVTCTYKAKIKELETILETQKVERSHSAKLEQDI 1147
LQEKNVTLDVQIQHVVEGKRALSELTVGVTCTYKAKIKELETILETQKVERSHSAKLEQDI
Sbjct: 61 LQEKNVTLDVQIQHVVEGKRALSELTVGVTCTYKAKIKELETILETQKVERSHSAKLEQDI 120

Query: 1148 LEKESIILKLEARNLKEFQEHLDQSVKNTKDLNVKELKLKEEITQLTNNLQDMKHLQLKE 1207
LEKESIILKLEARNLKEFQEHLDQSVKNTKDLNVKELKLKEEITQLTNNLQDMKHLQLKE
Sbjct: 121 LEKESIILKLEARNLKEFQEHLDQSVKNTKDLNVKELKLKEEITQLTNNLQDMKHLQLKE 180

Query: 1208 EEEETNRQETEKLEELSASSARTQNLKADLQRKEEDYADLKEKLTDAKKQIKQVQKEVS 1267
EEEETNRQETEKLEELSASSARTQNLKADLQRKEEDYADLKEKLTDAKKQIKQVQKEVS
Sbjct: 181 EEEETNRQETEKLEELSASSARTQNLKADLQRKEEDYADLKEKLTDAKKQIKQVQKEVS 240

Query: 1268 VMRDEDKLLRIKINELEKKKNQCSQELDMKQRTIQQLEQLNNQKVVEAIIQYERACKDL 1327
VMRDEDKLLRIKINELEKKKNQCSQELDMKQRTIQQLEQLNNQKVVEAIIQYERACKDL
Sbjct: 241 VMRDEDKLLRIKINELEKKKNQCSQELDMKQRTIQQLEQLNNQKVVEAIIQYERACKDL 300

Query: 1328 NVKEKIIEDMRMTLEEQQEQTQVEQDQVLEAKLEEVERLATELEKWKEKCNDETNNQRS 1387
NVKEKIIEDMRMTLEEQQEQTQVEQDQVLEAKLEEVERLATELEKWKEKCNDETNNQRS
Sbjct: 301 NVKEKIIEDMRMTLEEQQEQTQVEQDQVLEAKLEEVERLATELEKWKEKCNDETNNQRS 360

Query: 1388 NKEHENNTDVLGKLTNLQDELQSEQKYNADRKKWLEEKMLITQAKEAENIRNKEMKKY 1447
NKEHENNTDVLGKLTNLQDELQSEQKYNADRKKWLEEKMLITQAKEAENIRNKEMKKY
Sbjct: 361 NKEHENNTDVLGKLTNLQDELQSEQKYNADRKKWLEEKMLITQAKEAENIRNKEMKKY 420

Query: 1448 AEDRERFFKQONEMEILTAQLTEKDSDLQKWEERDQLVAALEIQLKALISSNVQKDNEI 1507
AEDRERFFKQONEMEILTAQLTEKDSDLQKWEERDQLVAALEIQLKALISSNVQKDNEI
Sbjct: 421 AEDRERFFKQONEMEILTAQLTEKDSDLQKWEERDQLVAALEIQLKALISSNVQKDNEI 480

Query: 1508 EQLKRIISETSKIETQIMDIKPKRISSADPKLQTEPLSTSFESISRNKIEDGSSVVLDSCE 1567
EQLKRIISETSKIETQIMDIKPKRISSADPKLQTEPLSTSFESISRNKIEDGSSVVLDSCE
Sbjct: 481 EQLKRIISETSKIETQIMDIKPKRISSADPKLQTEPLSTSFESISRNKIEDGSSVVLDSCE 540

Query: 1568 VSTENDQSTRFPKPELEIQFTPLQPNKMAVKHPGCTTPVTVEIPKARKRKSNEEEDLVK 1627
VSTENDQSTRFPKPELEIQFTPLQPNKMAVKHPGCTTPVTVEIPKARKRKSNEEEDLVK
Sbjct: 541 VSTENDQSTRFPKPELEIQFTPLQPNKMAVKHPGCTTPVTVEIPKARKRKSNEEEDLVK 600

Query: 1628 CENKKNATPRTNLKFPISDDRNSSVKKEQKVAIRPSSSKTYSLRSQASIIIGVNLATKKKE 1687
CENKKNATPRTNLKFPISDDRNSSVKKEQKVAIRPSSSKTYSLRSQASIIIGVNLATKKKE
Sbjct: 601 CENKKNATPRTNLKFPISDDRNSSVKKEQKVAIRPSSSKTYSLRSQASIIIGVNLATKKKE 660

Query: 1688 GTLQKFGDFLQHSPIQSILQSKAKKIETMSSSKLSNVEASKENVSQPKRAKRKLYTSEISS 1747
GTLQKFGDFLQHSPIQSILQSKAKKIETMSSSKLSNVEASKENVSQPKRAKRKLYTSEISS
Sbjct: 661 GTLQKFGDFLQHSPIQSILQSKAKKIETMSSSKLSNVEASKENVSQPKRAKRKLYTSEISS 720

Query: 1748 PIDISGQVILMDQKMKESDHQIIKRLRTKTAK 1780
PIDISGQVILMDQKMKESDHQIIKRLRTKTAK
Sbjct: 721 PIDISGQVILMDQKMKESDHQIIKRLRTKTAK 753

Score = 197 (29.6 bits), Expect = 2.1e-11, P = 2.1e-11
Identities = 114/542 (21%), Positives = 253/542 (46%)

Query: 692 IKTKLEELKKRENESDSLIQELETSSNKKIITQNRKIKELINIIDQKEDTINEFQNLKSHM- 750
+K + + E + I++L+ K +N R+KE + ++D + E + L +
Sbjct: 1 VKASSKKSHQIEELEQQIEKLQAEVKGKDYDENNRLEKEHE--KNQDDLLKEKETLIQQLK 58

Query: 751 ENTFFKCNDAKTS-SLIINNKLICNETVEVPKDSKSKICSERKRVNENELQQDEPPAK-- 807
E + N D ++ K +E + K+KI E + + E + + AK
Sbjct: 59 EELQEKNVTLDVQIQHVVEGKRALSELTVGVTCTYKAKI-KELETILETQKVERSHSAKLE 117

Query: 808 KGSIHVSSAITEDQKKSEEVPRNIAE-IEDIRVLQENNEGLRAFLLTIEENELKNEK---- 862

+ + S I + ++ +E + ++ + +++ + L L+ + + N L++ K
 Sbjct: 118 QDILEKESIIILKLERNLKEFQEHLDQSVKNTKDLNVKELKKEEITQLTNNLQDMKHLQ 177
 Query: 863 --EEKAELNKQIVH-FQELSLSEKKNLTLSKEVQIQSNYDIAIAELHVQSKNQEQEE 919
 EE+ E N+Q ++ELS S + L ++Q+ + +Y A+L K K + ++
 Sbjct: 178 LKEEEEETNRQETEKLEELSASSARTQNLKADLQRKEEDY----ADL---KEKLTDARK 230
 Query: 920 KIMKLSNEIETATRSITNNVSQIKLMHTKIDEL-RTLDVSVQISNIDLLNLRDLNSNGSEE 978
 +I ++ E+ S+ + + KL+ KI+EL + + SQ +D+ R + E+
 Sbjct: 231 QIKQVQKEV-----SVMRD--EDKLLRIKINELEKKNQCSQ--ELDMKQ-RTIQLKEQ 280
 Query: 979 DNLPTQLDLLGNDYLVSKQVKEYRIQEPNRENSFHSSIEAIWEECKEIVKASSKKSHQI 1038
 N N +++ Y + K+ ++E E+ ++E + E + K ++
 Sbjct: 281 LN--NQKVEEAIQQY--ERACKDLNVKEKIID-MRMTLEEQEQTQVEQDQVLEAKLEEV 335
 Query: 1039 EELEQQIEKLQAEVKGKDENNRLKEKEHKNQDDLLKEKETLIQQLKEELQEKVNT---- 1094
 E L ++EK + + + +NN+ KEH+N D+L + L +L+E Q+ N
 Sbjct: 336 ERLATELEKWKKECNDLETKNQSRNKEHENNTDVLGKLTNLQDELQSESEQYNADRKW 395
 Query: 1095 LDVQIQHVVEGKRA-----LSELTQGVTCYKAKIKELETILETQKVERSHSAKLEQDI 1147
 L+ ++ + + K A + + + + + + E+E IL Q E+ + ++
 Sbjct: 396 LEEKMMLITQAKEAENIRNKEMKKAEDRERFFKQONEME-ILTAQLTEKDSDLQKWR- 453
 Query: 1148 LEKESIIILKLERNLKEFQEHLDQSVKNTKDLNVKELK-LKEEITQLTNNLQDMKHLQLK 1206
 E++ ++ LE LK + +V+ KD +++LK + E +++ + D+K +
 Sbjct: 454 -ERDQLVALEIQLKAL---ISSNVQ--KDNEIEQLKRIISETSKIETQIMDIK---PKR 504
 Query: 1207 EEEEETNRQETEKLEELSASSARTQN 1233
 + ++ +TE L S + ++
 Sbjct: 505 ISSADPKLQTEPLSTSFEISRNKIED 531
 Score = 186 (27.9 bits), Expect = 3.2e-10, P = 3.2e-10
 Identities = 131/674 (19%), Positives = 294/674 (43%)
 Query: 673 LELKFNQIKAEAKTKGELIKT-KEELKKRENESDSLIELETSNKKIITONQRIKELIN 731
 L+ K ++ + +L K K LI+ KEEL+++ D IQ + + + Q +
 Sbjct: 35 LKEKEHKNQDDLLKEKETLIQQLKEELQEKVNTLDVQIQHVVEGKRALSELTOGVTCYKA 94
 Query: 732 IIDQREDTINEFQNL-KSHMENTFKCNDKADTSSLIINNKLICNETVEVPKDSKSKICSE 790
 I + E TI E Q + +SH + D + S+I+ + E E +DS
 Sbjct: 95 KIKELE-TILETQKVERSHSAKLEQ--DILEKESIIILKLERNLKEFQEHLDQSVKNTKDLNVKELKKEEITQLTNNLQDMKHLQLKKEEEEETNRQETEKLEELSASSARTQNL 207
 Query: 791 RKRNVNELQ-QDEPPAKKGSIHVSSAITEDQKKSEEV-RPNIAEI-EDIRVLQENNEGL 847
 K +N EL+ ++E ++ + + ++ EE R ++ E++ + L
 Sbjct: 148 TKDLNVKELKKEEITQLTNNLQDMKHLQLKKEEEEETNRQETEKLEELSASSARTQNL 207
 Query: 848 RAFLLTIENELKNEKEEKAELNKQIVHFQELSLSEKKNLTLSKEVQIQI-----QSNYDI 902
 +A L E + + KE+ + KQI Q+E+S+ ++ L ++ ++ Q + ++
 Sbjct: 208 KADLQRKEEDYADLKEKLTDAKKQIKQVQKEVSVMRDEDKLLRIKINELEKKNQCSQEL 267
 Query: 903 AIAELHVQSKNQEQEEKIMKLSNEIETATRSITNNVSQIKLMHTKIDEL-RTLDVSVQI 961
 ++ +Q+ +K Q +K+ + + E A + + I+ M ++E +T Q+
 Sbjct: 268 DMKQRTIQLKEQLNNQKVEEAIQQYERACKDLNVKEKIIDMRMTLEEQEQTQVEQDQV 327
 Query: 962 SNIDLLNLRDLNSNGSEEDNLPNTQLDLLGNDYLVSKQVKEYRI--QEPNRENSFHSSIEA 1019
 L + L+ E+ L+ N + + + N ++ S +
 Sbjct: 328 LEAKLEEERLATELEKWKKECNDLETKNQSRNKEHENNTDVLGKLTNLQDELQSESEQ 387
 Query: 1020 IWEECKEIVKASSKKSHQIEELEQQIEKLQAEVKGKDENNRLKEKEHKNQ--DDLLKEK 1077
 + K+ ++ Q +E E K E+K Y ++ R +++++ + L EK
 Sbjct: 388 YNADRKKWLEEKMLITQAKEAENIRNK---EMKKAEDRERFFKQONEMEILTAQLTEK 444
 Query: 1078 ETLIQQLKEELQEKVNTLDVQIQHVVEGKRALSELTOGVTCYKAKIKELETILETQKVER 1137
 ++ +Q+ +EE + L++Q++ ++ + + ++ ++ET + K +R
 Sbjct: 445 DSDLQKWRERDQLVALEIQLKALISSNVQKDNEIEQLKRIISETSKIETQIMDIKPKR 504
 Query: 1138 SHSAKLEQDILEKESIIILKLERNLKEFQEHLDQSVKNTKDLNVKELKKEEITQLT 1193
 SA ++ E S ++ RN E + DS +N + + +L+ + T L
 Sbjct: 505 ISSADPKLQTEPLSTSFEISRNKIEDGSSVLDSECVSTENDQSTRFPKPELEIQTPLQ 564
 Query: 1194 NNLQDMKH---LLQKKEEEEETNRQETEKLEEL-SASSARTQNLKADLQRKEEDYADLK 1249
 N +KH + + + +++++ ++E+L + + + +L+ D +
 Sbjct: 565 PNKMAVKHPGCTTPVTVKIPKARRKSNEMEEDLVKCNKKNATPRTNLKFPISSDRNSS 624
 Query: 1250 EKLTDARKQIKQVQKEVSVMRDEDKLLRIKINELEKKNQCSQEL-DMKQRTIQLKEQL 1308
 K + K I+ K+ +R + + I +N KKK Q+ D Q + L+ +
 Sbjct: 625 VK-KEQKVAIRPSSKKTYSLSRSQASI--IGVNLATKKKEGTLQKFGDFLQHSPIQLSKA 681
 Query: 1309 NNQKVEEAIQQYERACKDLNVKEKIIDMR 1338
 +K+ E + + + + + KE + + R
 Sbjct: 682 --KKIETMSSSKLSNVEAS-KENVSQPKR 708

Score = 165 (24.8 bits), Expect = 5.8e-08, P = 5.8e-08
Identities = 140/626 (22%), Positives = 271/626 (43%)

Query: 536 VEELENAEETQNVETKLLDEDLDKLTLEENKAFISHEEKRKLLDLIEDLKKKLINEKKEK- 594
+EELE E E K +D + L+E + H+ + LL E L ++L E +EK
Sbjct: 11 IEELEQQIEKLQAEVKGY-KDENNRLEKE-----HKNQDDLLKEKETLIQQLKEELQEK 65

Query: 595 LTLEFKIREEV-----QEFTQYWAQREADFKE--TLLQEREILEENAERRLAIFKDLVG 647
+TL+ +I+ V E TQ +A KE T+L+ +++ E + +L +D++
Sbjct: 66 VTLDVQIQHVVEGKRALSELTOGVTCYKAKIKELETILETQKV-ERSHSAKLE--QDILE 122

Query: 648 KCDT---REEAAKDICATKVETEEATACLELKFNQIKAEAKTKGELIKTKEELKKRENE 704
K E K+ ++ + T L +K ++K E+ + L K L+ +E E
Sbjct: 123 KESIILKLERNLKEFQEHLQDSVKNTKDLNVKELKLEETQLTNNLQDMKHLQLKEEE 182

Query: 705 SDSLIQELETSNKKIITQNRKIKELINIIDQKEDTINEFQNLKSHMENTFKCNDKADTSS 764
++ QE E +++ + R + L + +KE+ + + + K K + S
Sbjct: 183 EETNRQETEKLEELSASSARTQNLKADLQRKEEDYADLKEKLTDAKKQIKQVQK-EVSV 241

Query: 765 LIINNKLICNETVEVPKDSKSKICSERKRVNENELQQDEPPAKKGSIHVSSAITEDQKKS 824
+ +KL+ + E+ K K CS+ + + +QQ + V AI + ++
Sbjct: 242 MRDEDKLLRIKINELEK--KKNQCSQELDMKQRTIQQLKEQLNNQK--VEEAIQQYERAC 297

Query: 825 EEVRPNIAEIEDIRVLQENNEGLRAFLLTIEENELKNEKEEKAELNKQIVHFQQLSLSEK 884
+++ IEDR+ E E + + + L+ + EE L ++ +++ + E
Sbjct: 298 KDLNVKEKIIEDMRMTLEEQEQTQ---VEQDQVLEAKLEEVEERLATELEKWKECNDLET 354

Query: 885 KNLTLSEKVVQIQSNYDIAIAELHVQKSKNQEQEKEIMKLSNE-IETATRSITN-----N 938
KN S + + ++N D+ + +L + + QE E+K + +E IT N
Sbjct: 355 KNNQRSNK--EHENNTDV-LGKLTNLQDELQSEQKYNADRKKWLEEKMLITQAKEAEN 411

Query: 939 VSQIKLMHTKIDELRTLDSVSQISNIDL-LNLRD--LSNGSEEDNLPNTQLDLDLGNLYLV 995
+ ++ D R +++ + L +D L EE + L++ +
Sbjct: 412 IRNKEMKKAEDRERFFKQONEMEILTAQLTEKSDQLKWRREERDQLVALEIQLKALIS 471

Query: 996 SKQVKEYRIQEPNRENSFHSSIEA-IWE-ECKEIVKASSKSHQIEELEQQIEKLQAEVK 1053
S K+ I++ R S S IE I + + K I A K Q E L E + +++
Sbjct: 472 SNVQKDNEIEQLKRIISETSKIETQIMDIKPKRISSADPKL-QTEPLSTSFEISRNKIE 530

Query: 1054 GYKDNNRLKEKEHKNQDDLLKEKE-----TLIQQLKEELQEKNVTLDVQIQHVVEGKRA 1108
+ + +Q + E T +Q K ++ T V ++ KR
Sbjct: 531 DGSVVLDSCEVSTENDQSTRFPKPELEIQTPLQPNKMAVKHPGCTTPVTVKIPKARKRK 590

Query: 1109 LSELTOG-VTCYKAKIKELETILETQ-KVERSHSAKLEQDILEKES 1152
+E+ + V C K T L+ +R+ S K EQ + + S
Sbjct: 591 SNMEEDLVKCEKNKATPRTNLKFPISDDRNSSVKKEQKVAIRPS 636

Score = 143 (21.5 bits), Expect = 1.3e-05, P = 1.3e-05
Identities = 164/684 (23%), Positives = 304/684 (44%)

Query: 295 QKRKMLR-LSQDVKGYSFIKDLQWIVSDSKEAYRLKLGIKHQSVAFKLNASS---- 349
+K +++ L +++ + D+Q V + K A L G+ +L
Sbjct: 49 EKETLIQQLKEELQEKNVTLDVQIQHVVEGKRALSELTOGVTCYKAKIKELETILETQKV 108

Query: 350 -RSHSI-FTVKILQIEDSEMSRVIRVSELSLCDLAGSERTMKTQNEGE-RLRETGNINTS 406
RSHS IL+ E + + E L S + K N E +L+E T+
Sbjct: 109 ERSHSAKLEQDILEKESIILKLERNLKEFQE-HLQDSVKNTKDLNVKELKLEETQLTN 167

Query: 407 LTLGKCINVLNSEKSKFQQHVPFRESKLTHYFQSFFNGKGKICMIVNISQCYLAYDET 466
L K + LK E+ +Q + +L+ N K + + Y E
Sbjct: 168 NLQDMKHLQLKEEEEETNRQETEKLEELSASSARTQNLKADL---QRKEEDYADLKEK 224

Query: 467 LNVLFKSAIAQKVCVPTLNSSQDKLFGPVKSSQDVSLDSNSNSKILNVKRATISWENSL 526
L K I Q V ++ +DKL +K ++ + N S+ L++K+ TI
Sbjct: 225 LTDAK-KQIKQ-VQKEVSVMRDEKLLR-IKINE-LEKKKNQCSQELDMKQRTIQQLKEQ 280

Query: 527 EDLMEDELVEELENAEETQNVETKLLDEDLDKLTLEENKAFISHEEKRKLLDL-IEDLKK 585
+ + E+ +++ E A + NV+ K++ ED+ TLEE + + E+ ++L+ +E++++
Sbjct: 281 LNNQKVEEAIQQYERACKDLNVKEKII-EDMRMTLEEQEQTQ--TQVEQDQVLEAKLEEVEER 337

Query: 586 KLIN-EK-KEKLT-LEFKIREEVTOEFTQYWAQREADFKETLLQEREILEE-----NAERR 638
EK KEK LE K + +E + K T LQ+ E+ E NA+R+
Sbjct: 338 LATELEKWKECNDLETNNQRSNKEHEN---NTDVLGKLTNLQD-ELQSEQKYNADRK 393

Query: 639 LAIFKDLVGKCDTREAAKDICATKVETEEATACLELKFNQIKAEAKTKGELIKTKEEL 698
+ + ++ T+ + A++I K E ++ E F Q + E+ +L + +L
Sbjct: 394 KWLEEKMM--LITQAKEAENI-RNK-EMKKYAEDRERFFKQ-QNEMEILTAQLTEKSDSL 448

Query: 699 KKRENESDSLIQELETSNKKIITQN-QR---IKELINIIDQKEDTINEFQNLKSHMENTF 754
+K E D L+ LE K +I+ N Q+ I++L II + + ++K ++

Sbjct: 449 QKWREERDQLVAALEIQLKALISSNVQKDNEIEQLKRIISETSKIETQIMDIKPKRISSA 508

Query: 755 KCNDKADTSSLIINNKLICN--ETVEVPKDSKSKICSERK---RVNENELQ-QDEP--PA 806
DK T L + ++ N E V DS ++ +E R + EL+ Q P P

Sbjct: 509 D-PDKLQTEPLSTSFEISRNKIEDGSSVVLDS-CEVSTENDQSTRFPKPELEIQFTPLQPN 566

Query: 807 KKGSIH--VSSAITEDQKKSEEVPRNIAEIEDIRVLQENNEGLRA---FLLTIENELKNE 861
K H ++ +T K+ ++ N E + ++ + N R F ++ + +

Sbjct: 567 KMAVKHPGCTTPVTVKIPKARKRKSNEEEDLVKCNKKNATPRTNLKFPISDDRNSSVK 626

Query: 862 KEEKAEL---NKQIVHFQQLSLSEKKNLTLSKEVQIQSNYDIAIAELHVQSKNQEQE 918
KE+K + +K+ + + S+ NL K+ +Q D + +SK ++

Sbjct: 627 KEQKVAIRPSSKKTYSLSQASIIGV-NLATKKKEGTLQKFGDFLQHSFSLQSKAKKII 685

Query: 919 EKIM--KLSNEIETATRSITNNVSQIKLMHTKI--DELRT-LDSVQSISNID 965
E + KLSN +E + NVSQ K K+ E+ + +D Q+ +D

Sbjct: 686 ETMSSSKLSN-VEASKE----NVSQPKRAKRLYTSEISSPIDISGQVILMD 732

Score = 133 (20.0 bits), Expect = 1.6e-04, P = 1.6e-04
Identities = 94/426 (22%), Positives = 188/426 (44%)

Query: 527 EDLM-EDEDLVEELENAEETQNVETKLLDEDLDTLEENKAFISHEEKRLDL-IEDLK 584
+DL+ E E L+++L+ + +NV LD + +E +A + I++L+

Sbjct: 44 DDLLEKEETLIQQLKEELQEKNT---LDVQIQHVVEGKRALSELTQGVTCYKAKIKELE 100

Query: 585 KKLINKEKELTLEFKIREEVQ-EFTQYWAQREA-DFKETLLQEREILEENAERRLAIF 642
L +K E+ + K+ +++ + E +R +F+E L + ++ + L +

Sbjct: 101 TLETKQKVER-SHSAKLEQDILEKESIIKLERNLKEFQEHLDQSVKNTKDLNVKELKL- 158

Query: 643 KDLVGKCDTREAAKDICATKVETEEATACLELKFNQIKAEAKTKGELIKTKEELKKRE 702
K+ + + + K + K E EE + ++K EL+ + K +L+++E

Sbjct: 159 KEEITQLTNLQDMKHLQLKEEEETN---RQETEKLEELSASSARTQNLKADLQRKE 215

Query: 703 NESDSLQIELETSNKKIITONQRIKELINIIDQK-EDTINEFQNLKSHMENTFKCNDKA- 760
+ L ++L T KK I Q Q+ ++ D+ INE + K+ +

Sbjct: 216 EDYADLKEKL-TDAKKQIKQVQKESVVRDEDKLLRIKINELEKKKNQCSQELDMKQRTI 274

Query: 761 DTSSLIINNKLICNETVE---VPKDS--KSKICSE-RKRVNENE---LQQDEPPAKKGS 810
+NN+ + E ++ KD K KI + R + E E ++QD+ K

Sbjct: 275 QQLKEQLNNQKV-EEAIQYERACKDLNVKEKIIDEMRMTEEQEQTQVEQDQVLEAKLE 333

Query: 811 IHVSSAITEDQKKSEEVPR-NIAEIEDIRVLQENNEGLRAFLTIENELKNEKEEKAELN 869
V TE +K E+ + ENN + L +++EL+ E E+K +

Sbjct: 334 -EVLATELEKWKCKNDLETKNNQRSNKEHENNTDVLGKLTNLQDELQ-ESEQKYNAD 391

Query: 870 KQIVHFQQLSLSEKKNLTLSKEVQIQSNYDIAIAELHVQSKNQEQEEKIMKLSNEIE 929
++ ++++ L +T +KE + I++ + K E E+ K NE+E

Sbjct: 392 RK-KWLEEKMMML-----ITQAKEAENIRNK-----EMKKAEDRERFFKQQNEME 435

Query: 930 TATRSITNNVSQIKLMHTKIDEL 952
T +T S ++ + D+L

Sbjct: 436 ILTAQLTEKDSDLQKWREERDQL 458

Pedant information for DKF2phtes3_35b4, frame 3

Report for DKF2phtes3_35b4.3

[LENGTH] 1780
[MW] 206176.77
[pI] 5.60
[HOMOL] TREMBL:U93121_1 product: "M-phase phosphoprotein-1"; Human M-phase phosphoprotein-1 mRNA, partial cds. 0.0
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YEL061c] 2e-37
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YEL061c] 2e-37
[FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YEL061c] 2e-37
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YEL061c] 2e-37
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL058w] 7e-30
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDL058w] 7e-30
[FUNCAT] 30.05 organization of centrosome [S. cerevisiae, YPR141c] 3e-23
[FUNCAT] 11.01 stress response [S. cerevisiae, YPR141c] 3e-23
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YPR141c] 3e-23
[FUNCAT] 03.13 meiosis [S. cerevisiae, YPR141c] 3e-23
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YPR141c] 3e-23
[FUNCAT] 09.10 nuclear biogenesis [S. cerevisiae, YPR141c] 3e-23
[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YKR095w] 1e-21

[FUNCAT] 99 unclassified proteins [S. cerevisiae, YLR309c] 6e-20
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YHR023w
MYO1 - myosin-1 isoform] 4e-19
[FUNCAT] 03.25 cytokinesis [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 4e-19
[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YNL250w] 1e-15
[FUNCAT] 1 genome replication, transcription, recombination and repair [M.
jannaschii, MJ1322] 2e-14
[FUNCAT] 30.13 organization of chromosome structure [S. cerevisiae, YDR285w] 2e-09
[FUNCAT] 09.04 biogenesis of cytoskeleton [S. cerevisiae, YKL179c] 3e-09
[FUNCAT] 09.13 biogenesis of chromosome structure [S. cerevisiae, YLR086w] 2e-07
[FUNCAT] 03.01 cell growth [S. cerevisiae, YNL079c] 2e-07
[FUNCAT] 08.99 other intracellular-transport activities [S. cerevisiae, YNL079c]
2e-07
[FUNCAT] 03.22.01 cell cycle check point proteins [S. cerevisiae, YGL086w] 1e-06
[FUNCAT] 10.05.99 other pheromone response activities [S. cerevisiae, YHR158c]
3e-06
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YDR217c] 4e-06
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YJR134c] 2e-05
[FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae,
YAL035w] 2e-04
[FUNCAT] r general function prediction [M. jannaschii, MJ1254] 0.001
[BLOCKS] BL00387A
[BLOCKS] BL00411H
[BLOCKS] BL00411G
[BLOCKS] BL00411F
[BLOCKS] BL00411E Kinesin motor domain proteins
[BLOCKS] BL00411D Kinesin motor domain proteins
[BLOCKS] BL00411C Kinesin motor domain proteins
[BLOCKS] BL00411B Kinesin motor domain proteins
[BLOCKS] BL00411A Kinesin motor domain proteins
[SCOP] d2kin.1 3.29.1.5.3 Kinesin [Rat (Rattus norvegicus) 2e-68
[SCOP] d2tmab_ 1.105.4.1.1 Tropomyosin [rabbit (Oryctolagus cuniculus) 4e-05
[SCOP] d3kar_ 3.29.1.5.4 Kinesin [Baker's yeast (Saccharomyce 2e-09
[EC] 3.6.1.32 Myosin ATPase 5e-25
[PIRKW] nucleus 4e-27
[PIRKW] phosphotransferase 3e-16
[PIRKW] duplication 6e-20
[PIRKW] citrulline 6e-18
[PIRKW] tandem repeat 4e-24
[PIRKW] heterodimer 3e-28
[PIRKW] endocytosis 1e-23
[PIRKW] heart 1e-17
[PIRKW] transmembrane protein 2e-28
[PIRKW] serine/threonine-specific protein kinase 3e-16
[PIRKW] zinc finger 1e-23
[PIRKW] surface antigen 2e-16
[PIRKW] DNA binding 1e-25
[PIRKW] metal binding 1e-23
[PIRKW] muscle contraction 4e-24
[PIRKW] heterotetramer 4e-24
[PIRKW] acetylated amino end 2e-19
[PIRKW] actin binding 5e-25
[PIRKW] mitosis 3e-58
[PIRKW] microtubule binding 3e-58
[PIRKW] ATP 3e-58
[PIRKW] thick filament 4e-24
[PIRKW] phosphoprotein 9e-29
[PIRKW] leucine zipper 1e-12
[PIRKW] skeletal muscle 8e-24
[PIRKW] disulfide bond 1e-12
[PIRKW] heterotrimer 1e-29
[PIRKW] calcium binding 6e-18
[PIRKW] alternative splicing 4e-21
[PIRKW] P-loop 2e-63
[PIRKW] coiled coil 3e-58
[PIRKW] heptad repeat 1e-25
[PIRKW] methylated amino acid 4e-24
[PIRKW] peripheral membrane protein 1e-23
[PIRKW] dimer 1e-12
[PIRKW] cardiac muscle 1e-17
[PIRKW] hydrolase 5e-25
[PIRKW] microtubule 6e-15
[PIRKW] muscle 7e-23
[PIRKW] membrane protein 6e-20
[PIRKW] GTP binding 8e-22
[PIRKW] EF hand 6e-18
[PIRKW] cell division 1e-25
[PIRKW] cytoskeleton 4e-24
[PIRKW] hair 6e-18
[PIRKW] Golgi apparatus 8e-24
[PIRKW] calmodulin binding 1e-23

[SUPFAM] unassigned Ser/Thr or Tyr-specific protein kinases 3e-16
 [SUPFAM] myosin motor domain homology 5e-25
 [SUPFAM] alpha-actinin actin-binding domain homology 1e-13
 [SUPFAM] kinesin-related protein KIP1 9e-27
 [SUPFAM] kinesin-related protein CIN8 4e-36
 [SUPFAM] kinesin heavy chain 4e-24
 [SUPFAM] plectin 1e-13
 [SUPFAM] trichohyalin 6e-18
 [SUPFAM] kinesin-related protein KIF3 1e-29
 [SUPFAM] kinesin-related protein KIF2 3e-20
 [SUPFAM] ribosomal protein S10 homology 1e-13
 [SUPFAM] giantin 8e-24
 [SUPFAM] protein kinase homology 3e-16
 [SUPFAM] protein kinase C zinc-binding repeat homology 2e-13
 [SUPFAM] kinesin-related protein unc-104 8e-26
 [SUPFAM] human early endosome antigen 1 1e-23
 [SUPFAM] unassigned kinesin-related proteins 1e-28
 [SUPFAM] Mycoplasma genitalium hypothetical protein MG218 4e-17
 [SUPFAM] myosin heavy chain 5e-25
 [SUPFAM] conserved hypothetical P115 protein 4e-20
 [SUPFAM] centromere protein E 5e-24
 [SUPFAM] calmodulin repeat homology 6e-18
 [SUPFAM] kinesin-related protein KLP61F 1e-25
 [SUPFAM] hypothetical protein MJ0914 3e-12
 [SUPFAM] kinesin-related protein MKLP-1 2e-63
 [SUPFAM] pleckstrin repeat homology 8e-26
 [SUPFAM] hypothetical protein MJ1322 4e-13
 [SUPFAM] kinesin-related protein KIF1B 3e-28
 [SUPFAM] kinesin motor domain homology 2e-63
 [SUPFAM] kinesin-related protein KLPA 7e-25
 [SUPFAM] kinesin-related protein nodA 1e-12
 [SUPFAM] kinesin-related protein Eg5 5e-30
 [PROSITE] ATP_GTP_A 1
 [PFAM] Kinesin motor domain
 [KW] Irregular
 [KW] 3D
 [KW] LOW_COMPLEXITY 7.53 %
 [KW] COILED_COIL 19.78 %

SEQ MESNFNQEGVPRPSYVFSADPIARPSEINFDGIKLDLSHEFSLVAPNTEANSFESKDYLQ
 SEG
 COILS
 3kar-

SEQ VCLRIRPFTQSEKELESEGCVHILDSQTVVLKEPQCILGRLSEKSSGQMAQKFSFSKVFG
 SEG
 COILS
 3kar-

SEQ PATTQKEFFQGCIMQPVKDLLKGQSRLIFTYGLTNSGKTYTFQGTEENIGILPRTLNVLF
 SEG
 COILS
 3kar-

SEQ DSLQERLYTKMNLKPHRSREYLRLSSEQEKEEIASKALLRQIKEVTVHNSDDTLYGSL
 SEG
 COILS
 3kar-

SEQ TNSLNISEFEESIKDYEQANLNMANSIKFSVWVSFFEIYNEYIYDLFVPVSSKFQKRKML
 SEG
 COILS
 3kar-EEEEEEEEETTEEEETTTC-----CCEE

SEQ RLSQDVKGYSFIKDLQWIQVSDSKEAYRLKLGKHSVAFTKLNNASSRSHSIFTVKIL
 SEG
 COILS
 3kar- EEETTTTE-EEETTTCCEEECCGGHHHHHHHHHHHCCCTTTTCHHHHHHCEEEEEEEEE

SEQ QIEDSEMSRVIRVSELSLCLAGSERTMKTQNEGERLRETGNINTSLTLGKNCINVLKNS
 SEG
 COILS
 3kar- E--EETTTTCEEEEEEEEECCCCCCC---CCCHHHHHHHHHHHHHHHHHHHHHHTT

SEQ EKSQFQHVFPRESKLTHTYFQSFNGKGIKICMIVNISQCYLAYDETLNVLFSAIAQKVC
 SEG
 COILS
 3kar- TTTT--TCCTTTTTHHHHHHGGGCTTTTEEEEEEECCGGHHHHHHHHHHH.....

SEQ VPDTLNSSQDKLFGPVKSSQDVSLDSNSKILNVKRATISWENSLEDLMEDEDLVEELE

SEGXXXXXXXXXXXXXXXXXXXX
COILS
3kar-

SEQ NAEETQNVETKLLDEDLDKTLLEENKAFISHEEKRLDLIEDLKKKLINEKKEKLTLEFK
SEGXXXXXXXXXXXXXXXXXXXX
COILSCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
3kar-

SEQ IREEVTQEFQYWAQREADFKETLLQEREILEENAERRLAIFKDLVGKCDTREAAKDIC
SEG
COILS CCCCCC.....
3kar-

SEQ ATKVETEEATACLELKFNQIKAEAKTKGELIKTKEELKKRENESDSLIELETSNKKII
SEG
COILSCCCCCCCCCCCCCCCC
3kar-

SEQ TQNRIKELINIIDQKEDTINEFQNLKSHMENTFKCNDKADTSSLIINNKLICNETVEVP
SEG
COILS CCCCCCCCCCCCCC.....
3kar-

SEQ KDSKSKICSERKRVNENELQQDEPPAKKGSIHVSSAITEDQKKSEEVNPNIAEIEDIRVL
SEG
COILSCCCC
3kar-

SEQ QENNEGLRAFLTLTIENELKNEKEEKAELNKQIVHFQOELSSEKKNLTLSKEVQOIQSNY
SEGXXXXXXXXXXXXXXXXXXXX
COILS CC.....
3kar-

SEQ DIAIAELHVQSKSNQEQEEKIMKLSNEIETATRSITNNVSIKLMHTKIDELRTLDSVSQ
SEG
COILSCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....
3kar-

SEQ ISNIDLLNLRDLSNGSEEDNLPNTQLDLLGNDYLVSKQVKEYRIQEPNRENSFHSSIEAI
SEG
COILS
3kar-

SEQ WEECKEIVKASSKKSHQIEELEQQIEKLQAEVKGKDNENRLKEKEHKNQDDLLKEKETL
SEGXXXXXXXXXXXXXXXXXXXX
COILSCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....
3kar-

SEQ IQQLKEELQEKVNTLDVQIQHVVEGKRALSELTOGVTCYKAKIKELETILETQKVERSUS
SEG
COILSCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....
3kar-

SEQ AKLEQDILEKESIILKLERNLKEFQEHLQDSVKNTKDLNVKELKLKEITQLTNNLQDMK
SEG
COILS CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....CCCCCCCCCCCCCCCCCCCCCCCCCCCC
3kar-

SEQ HLLQLKEEEEETNRQETEKLEELSASSARTQNLKADLQKEEDYADLKEKLTDAKKQIK
SEGXXXXXXXXXXXXXXXXXXXX
COILS CCCCC.....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
3kar-

SEQ QVQKEVSVMRDEKLLRIKINELEKKKNQCSQELDMKQRTIQQLKEQLNNQKVEEAIQOY
SEG
COILS CCCCCCCCCC.....
3kar-

SEQ ERACKDLNVKEKIIEDMRMTLEEQEQTQVEQDQVLEAKLEEVERLATELEKWKKEKCNLE
SEGXXXXXXXXXXXXXXXXXXXX
COILSCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
3kar-

SEQ TKNNQRSNKEHENNTDVLGKLTNLQDELQEQKYNADRKKWLEEKMMITQAKEAENIR
SEG
COILS CC.....
3kar-

SEQ NKEMKKYAEADRERFFKQONEMEILTAQLTEKDSDLQKWREERDQLVAALEIQLKALISSN
SEG

```

COILS .....
3kar- .....

SEQ      VQKDNEIEQLKRIISETSKIETQIMDIKPKRISSADPKLQTEPLSTSFEISRNKIEDGS
SEG      .....
COILS    .....
3kar-    .....

SEQ      VVLDSCEVSTENDQSTRFPKPELEIQFTPLQPNKMAVKHPGCTTPVTVEIPKARKRSNE
SEG      .....
COILS    .....
3kar-    .....

SEQ      MEEDLVKCNKKKNATPRTNLKFPISDDRNSSVKKEQKVAIRPSSKKTYSLSQASIIGVN
SEG      .....
COILS    .....
3kar-    .....

SEQ      LATKKKEGTLQKFGDFLQHSPIQS KAKKIIETMSSSKLSNVEASKENVSQPKRAKRKL
SEG      .....
COILS    .....
3kar-    .....

SEQ      YTSEISSPIDISGQVILMDQKMKESDHQIIKRRRLRTKTAK
SEG      .....
COILS    .....
3kar-    .....

```

Prosites for DKFZphtes3_35b4.3

PS00017 152->160 ATP_GTP_A PDOC00017

Pfam for DKFZphtes3_35b4.3

```

HMM_NAME      Kinesin motor domain

HMM            *RCRPlNeEindgcscvVQWPpWtGyktvhngheds.....phks
               R+RP+ + E++ + +V + +++ ++ + ++
Query         64  RIRPFTQSEKELESEGCVHILDQSTVVLKEPQCILGRLSEKSSGQMAQK  112

HMM            FtFDHVFVWncTQedVYdtvAHPIVDDcFhGYNCTIFAYGQTGSGKTYTM
               F+F +VF++++TQ++ +++ + V+D+++G IF+YG T SGKTYT
Query         113 FSFSKVFGPATTQKEFFQGCIMQPVKDLLKGQSRLIFTYGLTNSGKTYTF  162

HMM            MGpggehPDHmGIIPRcCHDIFdrIdkfqekDhdFW.....
               G +++GI+PR+++ +FD++ + +++
Query         163 QG----TEENIGILPRTLNVLFDSLQERL-YTKMNLKPHRSREYLRLSSE  207

HMM            .....
Query         208 QEKEEIASKSALLRQIKEVTVHNDDTLTGLTNSLNISEFEESIKDYE  257

HMM            .....hVkcSYMEIYNEeIYDLLCPnP...qhMkpLnIHEHPN
               +V +S++EIYNE+IYDL +P++ Q++K L++ + +
Query         258 QANLNMANSIKFSVWVSFFEIYNEYIYDLFVPVSSKFQKRKMLRLSQDVK  307

HMM            MGpYVqGCTEfHVCsYeDachWIWqGnknRHVAaTnMndhSSRShtIFTI
               ++++++ V +A +++ +G K+ VA T++N SSRShtIFT+
Query         308 GYSFIKDLQWIQVSDSKEAYRLLKLGIKHQSVAFTKLNNASSRShtIFTV  357

HMM            HVeQrHk.qcdehvcHskMNLVDLAGSERvnrTGAEGQRlKEGcNINqSL
               ++ Q + + +++S ++L DLAGSER+ +T+ EG RL+E +NIN SL
Query         358 KILQIEDSEMSRVIRVSELSLCDLAGSERTMKTQNEGERLRETGNINTSL  407

HMM            ttLGnVInaLaDgqTKYmYgghgHIPYRDSKLTWlLQDSLGGNcKtCMIA
               +TLG++IN+L + + + +H+P+R+SKLT+ +Q + G +K CMI+
Query         408 LTLGKCINVLKNS---KSKFQQHVFFRESKLTHTYFQSFFNGKGKICMIV  454

HMM            CIWPadWNYEETLSTLRYAdRAKnIkNkPQINEDPca*
               +I+ + Y+ETL++L++ + A+++ + ++N+++++
Query         455 NISQCYLAYDETLNVLFKSAIAQKVCVPDTLNSSQDK  491

```

DKFZphtes3_35b5

group: metabolism

DKFZphtes3_35b5 encodes a novel 466 amino acid protein, with similarity to bovine accessory subunit for vacuolar ATPase and rat C7-1 protein.

The vacuolar proton-ATPase (V-ATPase) translocates protons into intracellular organelles or across the plasma membrane of specialized cells. The catalytic domain consists of a hexamer of 3 A subunits and 3 B subunits, plus accessory subunits C, D, and E. The rat homolog C7-1 seems to be enriched in aged adult rats in the frontal cortex.

The novel protein can find application in modulating the v-ATPase activity in endocytic and secretory organelles.

strong similarity to bovine vacuolar ATPase (EC 3.6.1.-) chain A

complete cDNA, complete cds potential start at Bp 8 , EST hits matches perfect to I54197 hypothetical protein, but possess 186 aa additional at N-terminus

Sequenced by DKFZ

Locus: unknown

Insert length: 2043 bp

Poly A stretch at pos. 2033, polyadenylation signal at pos. 2012

```
1  GCGGCCCATG GCGACGGCTC GAGTGC GGAT GGGGCCGCGG TGC GCCCAGG
51  CGCTCTGGCG CATGCCGTGG CTGCCGGTGT TTTTGTCTGT GCGGCCGCGG
101 GCGGCCGCGG CAGCGGCGGA GCAGCAGGTC CCGCTGGTGC TGTGGTCGAG
151 TGACCGGGAC TTGTGGGCTC CTGCCGCCGA CACTCATGAA GGCCACATCA
201 CCAGCGCACT GCAGCTCTCT ACCTACTTAG ATCCCGCCCT GGAGCTGGGT
251 CCCAGGAATG TGTGCTGTGT CCTGCAGGAC AAGCTGAGCA TTGAGGATTT
301 CACAGCATAT GCGGGTGTGT TTGGAACAA GCAGGACAGC GCCTTTTCTA
351 ACCTAGAGAA TGCCCTGGAC CTGGCCCCCT CCTCACTGGT GCTTCCTGCC
401 GTCGACTGGT ATGCAGTCAG CACTCTGACC ACTTACCTGC AGGAGAAGCT
451 CGGGGCCAGC CCCTTGCA TGACCTGGC CACCCTGCGG GAGCTGAAGC
501 TCAATGCCAG CCTCCCTGCT CTGCTGCTCA TTCGCCTGCC CTACACAGCC
551 AGCTCTGGTC TGATGGCACC CAGGGAAGTC CTCACAGGCA ACGATGAGGT
601 CATCGGGCAG GTCCTGAGCA CACTCAAGTC CGAAGATGTC CCATACACAG
651 CGGCCCTCAC AGCGGTCCGC CCTTCCAGGG TGGCCCGTGA TGTAGCCGTG
701 GTGGCCGGAG GGCTAGGTCG CCAGCTGCTA CAAAAACAGC CAGTATCACC
751 TGTGATCCAT CCTCTGTGTA GTTACAATGA CACCGCTCCC CGGATCCTGT
801 TCTGGGCCCA AAACCTCTCT GTGGCGTACA AGGACCAGTG GGAGGACCTG
851 ACTCCCTCA CCTTTGGGGT GCAGGAAGTC AACCTGACTG GCTCCTTCTG
901 GAATGACTCC TTGCCCAGGC TCTCACTGAC CTATGAACGA CTCTTTGGTA
951 CCACAGTGAC ATTCAAGTTC ATTCTGGCCA ACCGCCTCTA CCCAGTGTCT
1001 GCGCGGCACT GGTTTACCAT GGAGCGCCTC GAAGTCCACA GCAATGGCTC
1051 CGTCGCCTAC TTCAATGCTT CCCAGGTCAC AGGGCCAGC ATCTACTCCT
1101 TCCACTGCGA GTATGTCAGC AGCCTGAGCA AGAAGGGTAG TCTCCTCGTG
1151 GCGCGCACGC AGCCCTCTCC CTGGCAGATG ATGCTTCAGG ACTTCCAGAT
1201 CCAGGCTTTC AACGTAATGG GGGAGCAGTT CTCCTACGCC AGCGACTGTG
1251 CCAGCTTCTT CTCCCCCGGC ATCTGGATGG GGCTGCTCAC CTCCCTGTTT
1301 ATGCTCTTCA TCTTCACTA TGGCCTGCAC ATGATCCTCA GCCTCAAGAC
1351 CATGGATCGC TTTGATGACC ACAAGGGCCC CACTATTCTT TTGACCCAGA
1401 TTGTGTGACC CTGTGCCAGT GGGGGGGTTG AGGGTGGGAC GGTGTCCGTG
1451 TTGTTGCTTT CCCACCCTGC AGCGCACTGG ACTGAAGAGC TTCCCTCTTC
1501 CTACTGCAGC ATGAAGTGCA AGCTCCCTC AGCCCATCTT GCTCCCTCTT
1551 CAGCCCGCTG AGGAGCTTTC TTGGGCTGCC CCCATCTCTC CCAACAAGGT
1601 GTACATATTC TGCGTAGATG CTAGACCAAC CAGCTTCCCA GGGTTCGTCG
1651 CTGTGAGGCG TAAGGGACAT GAATTCTAGG GTCTCCTTTC TCCTTATTTA
1701 TTCTTGTGGC TACATCATCC CTGGCTGTGG ATAGTGCTTT TGTGTAGCAA
1751 ATGCTCCCTC CTTAAGGTTA TAGGGCTCCC TGAGTTTGGG AGTGTGGAAG
1801 TACTACTTAA CTGTCTGTCC TGCTTGGCTG CCGTTATCGT TTTCTGGTGA
1851 TGTGTGCTA ACAATAAGAA GTACACGGGT TTATTCTCTG GGCCTGAGAA
1901 GGAAGGGACC TCCACGACAG GTGGGCTGGG TCGATCGGCC GGCTGTTTGG
1951 CATGTTCCCA CCGGGAGTGC CGGGCAGGAG CATGGGGTGC TTGGTTGTTT
2001 CCTTCTAAT AAAATAAAGC CGGGTGCCA TGCAAAAAA AAA
```

BLAST Results

No BLAST result

Medline entries

95014142:

A novel accessory subunit for vacuolar H(+)-ATPase from chromaffin granules.

97215246:

Identification of a rat brain gene associated with aging by PCR differential display method.

Peptide information for frame 2

ORF from 8 bp to 1405 bp; peptide length: 466
Category: strong similarity to known protein

```

1 MATARVRMGP RCAQALWRMP WLPVFLSLAA AAAAAAAEQQ VPLVLWSSDR
51 DLWAPAADTH EGHITSDLQL STYLDPALEL GPRNVLLFLQ DKLSIEDFTA
101 YGGVFGNKQD SAFSNLENAL DLAPSSLVLP AVDWYAVSTL TTYLQEKLG
151 SPLHVDLATL RELKLNALSL ALLLIRLPYT ASSGLMAPRE VLTGNDEVIG
201 QVLSTLKSSED VPYTAALTAV RPSRVARDVA VVAGGLGRQL LQKQPVSPVI
251 HPPVSYNDTA PRILFWAQN FSVAYKDQWED LTPLTFGVQE LNLGSEFWND
301 SFARLSLTYE RLFGTTVTFK FILANRLYPV SARHWTMER LEVHNSGVA
351 YFNASQVTGP SIYSFHCEYV SLSKKGSL VARTQSPWQ MMLQDFQIQ
401 FNVMEQFSY ASDCASFFSP GIWMGLTSL FMLFIFTYGL HMILSLKTMD
451 RFDDHKGPTI SLTQIV

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_35b5, frame 2

TREMBL:AF035387.1 gene: "C7-1"; product: "C7-1 protein"; Rattus norvegicus C7-1 protein (C7-1) mRNA, complete cds., N = 1, Score = 2088, P = 3.8e-216

PIR:A55116 vacuolar ATPase (EC 3.6.1.-) chain Ac45 - bovine, N = 1, Score = 2011, P = 5.5e-208

PIR:I54197 hypothetical protein - human, N = 1, Score = 1464, P = 5.1e-150

>TREMBL:AF035387.1 gene: "C7-1"; product: "C7-1 protein"; Rattus norvegicus C7-1 protein (C7-1) mRNA, complete cds.
Length = 463

HSPs:

Score = 2088 (313.3 bits), Expect = 3.8e-216, P = 3.8e-216
Identities = 408/463 (88%), Positives = 426/463 (92%)

```

Query:      4 ARVRMGPRCAQALWRMPWLPVFLSLAAAAAAEQQVPLVLWSSDRDLWAPADTHEGH 63
             +R+R G R A  LW      + LSL A AAA AAEQQVPLVLWSSDRDLWAP ADTHEGH
Sbjct:      8 SRIRTGTRWAPVLW-----LLLSLVAVAAVAEQQVPLVLWSSDRDLWAPVADTHEGH 61

Query:     64 ITSDLQLSTYLDPALELGPRNVLLFLQDKLSIEDFTAYGGVFGNKQDSAFSNLENALDLA 123
             ITSD+QLSTYLDPALELGPRNVLLFLQDKLSIEDFTAYGGVFGNKQDSAFSNLENALDLA
Sbjct:     62 ITSDMQLSTYLDPALELGPRNVLLFLQDKLSIEDFTAYGGVFGNKQDSAFSNLENALDLA 121

Query:     124 PSSLVLPVADWYAVSTLTYYLQEKLGASPLHVDLATLRELKLNALSLPALLLIRLPYTASS 183
             PSSLVLPVADWYA+STLTYYLQEKLGASPLHVDLATL+ELKLNALSLPALLLIRLPYTASS
Sbjct:     122 PSSLVLPVADWYAISTLTYYLQEKLGASPLHVDLATLRELKLNALSLPALLLIRLPYTASS 181

Query:     184 GLMAPREVLTGNDEVIGQVLSTLKSSEDVPYTAALTAVRPSRVARDVAVVAGGLGRQLLQK 243
             GLMAPREVLTGNDEVIGQVLSTL+SEDPYTAALTAVRPSRVARDVA+VAGGLGRQLLQ
Sbjct:     182 GLMAPREVLTGNDEVIGQVLSTLESEDPYTAALTAVRPSRVARDVAMVAGGLGRQLLQT 241

Query:     244 QPVSPVHPPVSYNDTAPRILFWAQNFSVAYKDQWEDLTPLTFGVQELNLTGSFWNDSFA 303
             Q  SP IHPPVSYNDTAPRILFWAQNFSVAYKD+W+DLT LTFGV+ LNLGSEFWNDSFA
Sbjct:     242 QVASPAPHPPVSYNDTAPRILFWAQNFSVAYKDEWKDLTSLTFGVENLNLGSEFWNDSFA 301

Query:     304 RLSLTIERLFGTTVTFKFILANRLYPVSARHWTMERLEVHNSGVSVAFYFNASQVTGPSIY 363
             LSLTYE LFG TVTFKFILAN+R YPVSA+RWTMERLE+HSNGSVA+FN SQVTGPSIY

```


Sbjct: 302 MSLTYEPLFGATVTFKFIASRFYPVSARYWFTMERLEIHSNGSVAHFNVSQVTGPSIY 361
 Query: 364 SFHCEYVSSLSKKGSLLVARTQPSWQMLQDFQIQAFNVMGEOFSYASDCASFFSPGIW 423
 SFHCEYVSSLSKKGSLLV PS WQM L +FQIQAFNV GEQFSYASDCA FFSPGIW
 Sbjct: 362 SFHCEYVSSLSKKGSLLVNTV-PSLWQMTLHNFQIQAFNVTEQFSYASDCAGFFSPGIW 420
 Query: 424 MGLLTSLFMLFIFTYGLHMILSLKTMDFDDHKGPTISLTQIV 466
 MGLLT+LFMLFIFTYGLHMILSLKTMDFDD KGPTI+LTQIV
 Sbjct: 421 MGLLTSLFMLFIFTYGLHMILSLKTMDFDDHKGPTITLTQIV 463

Pedant information for DKFZphtes3_35b5, frame 2

Report for DKFZphtes3_35b5.2

[LENGTH] 466
 [MW] 51621.44
 [pI] 5.73
 [HOMOL] TREMBL:AF035387_1 gene: "C7-1"; product: "C7-1 protein"; Rattus norvegicus C7-1
 protein (C7-1) mRNA, complete cds. 0.0
 [PIRKW] hydrolase 0.0
 [PROSITE] MYRISTYL 7
 [PROSITE] CAMP_PHOSPHO_SITE 1
 [PROSITE] CK2_PHOSPHO_SITE 7
 [PROSITE] TYR_PHOSPHO_SITE 1
 [PROSITE] PKC_PHOSPHO_SITE 8
 [PROSITE] ASN_GLYCOSYLATION 7
 [KW] SIGNAL_PEPTIDE 38
 [KW] TRANSMEMBRANE 1
 [KW] LOW_COMPLEXITY 11.59 %

SEQ MATARVRMGPRCAQALWRMPWLPVFLSLAAAAAAAAAEQQVPLVLWSSDRDLWAPAADTH
 SEGxxxxxxxxx.....
 PRD cccccccccchhhhhhhccccchhhhhhhhhhhhhhhcccccccccccccccccc
 MEM
 SEQ EGHITSDQLSTYLDPALELGPRNVLLFLQDKLSIEDFTAYGGVFGNKQDSAFSNLENAL
 SEG
 PRD cccccchhhhhccccccccccccccccccccccccccccccccccccchhhhhhhhhcc
 MEM
 SEQ DLAPSSLVLPADVWYAVSTLTLYLQEKLGASPLHVDLATLRELKLNASLPALLLIRLPYT
 SEGxxxxxxxxxxxxxxxxx.....
 PRD ccc
 MEM
 SEQ ASSGLMAPREVLGNDEVIGQVLSTLKSIEDVPYTAALTAVRPSRVARDVAVVAGGLGRQL
 SEGxxxxxxxxxxxxxxxxxxxxxxxx.....
 PRD ccc
 MEM
 SEQ LQKQPVSPVIHPPVSYNDTAPRILFWAQNFSVAYKQWEDLTPLTFGVQELNLTGSFWND
 SEG
 PRD hhhhhcc
 MEM
 SEQ SFARLSLTYERLFGTTVTFKFIANRLYPVSARHWFTMERLEVHNSGSAVFNASQVTGP
 SEG
 PRD hhhhhhhhhhhhhcc
 MEM
 SEQ SIYSFHCEYVSSLSKKGSLLVARTQPSWQMLQDFQIQAFNVMGEOFSYASDCASFFSP
 SEGxxxxxxxxx.....
 PRD ccc
 MEMMMMMM
 SEQ GIWMGLLTSLFMLFIFTYGLHMILSLKTMDFDDHKGPTISLTQIV
 SEG
 PRD ccchhhcccccccccccccc
 MEM MM

Prosite for DKFZphtes3_35b5.2

PS00001	166->170	ASN_GLYCOSYLATION	PDOC00001
PS00001	257->261	ASN_GLYCOSYLATION	PDOC00001
PS00001	269->273	ASN_GLYCOSYLATION	PDOC00001

PS00001	292->296	ASN_GLYCOSYLATION	PDOC00001
PS00001	299->303	ASN_GLYCOSYLATION	PDOC00001
PS00001	346->350	ASN_GLYCOSYLATION	PDOC00001
PS00001	353->357	ASN_GLYCOSYLATION	PDOC00001
PS00004	375->379	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	3->6	PKC_PHOSPHO_SITE	PDOC00005
PS00005	48->51	PKC_PHOSPHO_SITE	PDOC00005
PS00005	159->162	PKC_PHOSPHO_SITE	PDOC00005
PS00005	205->208	PKC_PHOSPHO_SITE	PDOC00005
PS00005	318->321	PKC_PHOSPHO_SITE	PDOC00005
PS00005	331->334	PKC_PHOSPHO_SITE	PDOC00005
PS00005	374->377	PKC_PHOSPHO_SITE	PDOC00005
PS00005	445->448	PKC_PHOSPHO_SITE	PDOC00005
PS00006	48->52	CK2_PHOSPHO_SITE	PDOC00006
PS00006	72->76	CK2_PHOSPHO_SITE	PDOC00006
PS00006	94->98	CK2_PHOSPHO_SITE	PDOC00006
PS00006	114->118	CK2_PHOSPHO_SITE	PDOC00006
PS00006	159->163	CK2_PHOSPHO_SITE	PDOC00006
PS00006	193->197	CK2_PHOSPHO_SITE	PDOC00006
PS00006	255->259	CK2_PHOSPHO_SITE	PDOC00006
PS00007	207->214	TYR_PHOSPHO_SITE	PDOC00007
PS00008	102->108	MYRISTYL	PDOC00008
PS00008	103->109	MYRISTYL	PDOC00008
PS00008	200->206	MYRISTYL	PDOC00008
PS00008	295->301	MYRISTYL	PDOC00008
PS00008	314->320	MYRISTYL	PDOC00008
PS00008	421->427	MYRISTYL	PDOC00008
PS00008	425->431	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_35b5.2)

DKF2phtes3_35e21

group: differentiation/development

DKF2phtes3_35e21.2 encodes a novel 104 amino acid putative interleukin precursor, related to interleukin-7.

Due to the close relationship to human interleukin-7, the novel interleukin is expected to act as a new growth factor for human B lineage cells. Additionally, the protein should induce the gene rearrangement of the T-cell receptor repertoire, leading to thymocyte commitment, and subsequently induce both cytotoxic T-cell- and lymphocyte-activated killer cells.

This new interleukin could find clinical application in a variety of conditions of hematolymphopietic failure and different tumours, because of its recruitment of B cell lineage cells, cytotoxic T-cell- and lymphocyte-activated killer cells.

similarity to interleukin-7 precursor

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 2095 bp

Poly A stretch at pos. 2085, polyadenylation signal at pos. 2067

```
1 GGATGAAAGT GATTTAATTC ATTTTGTAGAA TTTTGTGTTT GTTTTGTGTTT
51 AGCAACATGC TGAACAACCTA ATTTACTTTA AAAATAAGCC AGTTAAAAACA
101 AAGGACGCTA AGCCCAAGTG GGGGGCAATA TTAGTCAGGA TCTTTGGGGT
151 CTAATTCAGC ACCAAGCTTC AGAAGCAGTT CTTTGTCTCT GTTCTCACCT
201 CTGCTGTCCC TCTCTTCCCT CATCCCCTAA GAGAGACAAA GATAAAAGCC
251 CACCTGCATC CCTAAGTCTT ACTGAGATCA GCCACCCAG GGGAGAGAAA
301 CTGGATCTAC TTACAGCCAC CCCCTGTTTC CATCCATATA CTTACTTCCC
351 CCAATTTGCA TGTGATTATG GAAACAAGTC ATGCTCATGA AAGCAACTGT
401 AAAATAAAGG GTTATGGAGT AGTTCAGCAA CTTCTTCACA GCCAGCTTTG
451 TGGAGCTGGG GAGGACTTAG GGCCCATGG AGTCTCTTAT GTGTACAGCT
501 TCAGGGCTGT CCCTTTCAGT TTGATTTTAA GCAATGCCTC ACTTCATAGC
551 TTAGGGGGTA AGGATTCAT TCAGGTAGGT TGTCTAAAGG AACTAATGGG
601 ACCTCTCAGT GAATTAGCTG ACCAGATTTT AGGAAATCTT TTTAATTTCT
651 ATGATTTTCC TTCTCAGATT TTGAAATGGT AAAATGACTT GGAAATAATT
701 TTTCTTGGTG CCTTATTGGT TTTCTTGAAC AACCTTCTC ATATTTTCTC
751 ATGACCAATG CCAGTGACCA AGGCCCATGT GTGTGTTGTG TGTAAATGTG
801 GGCATGTACA AGCTTAAATA ACGTGCCGAC AGCACTGTTT CAAAGTTGGT
851 ATTCATTAGG CTGTTGCCTC CTGGGCTGGA GCTGCGCTAA TCCTGACACC
901 GGTTGCCAGG AGAAAACCTC ATGGATCACA CACCAACACG TAATAACAGC
951 ATCCGTGACC TGCACCTCC AGTACAGAA GGGAAACCCA GAGCTAGGAA
1001 ATGTAGTTGT ATATTTTAAT GAAGTGTAC CCCAGCCAAA GAAGCTTCTT
1051 TCACTTTTGT GCTCTACAGA AAGCCCAAGG GGGGTAGGAG GGACAGAGCT
1101 TTGAATAACT GCTTTCTAAC ACTAAATGTG GCCAACAGGA CAGAGCACAT
1151 CACAGCTGTA GGCAGGTGTG AGGGACAGTG GCTAAGAATT GCCTGCTCCC
1201 TCTGCATGCT CTTTCTTGT TCCAAAGTCC AATCAAGTGA TCCTGGGAAA
1251 CAAATCTGTC TGGATTGCGG AGGGTGGTTC TGAAAGAACT GCCAAGACGT
1301 TAAAGAAAGG TGAAGAGTAG GCAGAATATA AGTAGCTAAC CTGAGTCAAG
1351 ACTCTCAAAA GCTAGCAGCC TGATGACAAT AGGATTTATT TCAGCCAGGA
1401 TAGTGCTGCT CTGTGAGTGC ATCATTTTAA GACAGTATGA CTTCAATGTT
1451 TTACAAACTA TGTATAGTAT GTATGTTTGT TGGGTTGTAT ATATACATAA
1501 TATATATTAT ATATATATAT GAGAGATTGT GTGACTTTTG ATACGGGTTT
1551 GGTGCAGGTG AATTTATTAC TGAGCCAAAT GAGGCACATA CCGAGTCAGT
1601 AGTTGAAGTC CAGGGCATTC GATACTGTTT ATGATTTCCTA TATATGTATA
1651 GTGCTATCC CATGCTGTAG TCACTGTTAT GTTAAATCCA GAAGTACAC
1701 TAGAGCCAGC GATACTTTAT TTGTAGACAA TCAATTTGAA TCCATATGTT
1751 ATTACTGGCA GATGATACAT GATTACAGTT CTGAATCTGT AACACTTACA
1801 AAAGGAAACC CAGAGCAGCT TGATGAGTTT TTGTTTCTGC TTCGTTCCCTG
1851 GGAGTCAGTA GAAACAGCAG TTGTATGTGG TTATGTTAGT CTCAGATAC
1901 TTAATTTGTT GACCTTACTT CAGAAAAATT TTGTATGTAT TATATTTGTG
1951 GGAAGGTAAT ATAATCATTT GAGATTTTGA TCAAAATATGA AGATTAGTTA
2001 TTTATGAAAA ACAAGAAAT GTCTATTTT CTTGTTCCTC AATTAATGTA
2051 GATAAATTTT AAAATGCATT AAAGTAATGG TCCGGAAAAA AAAAA
```

BLAST Results

No BLAST result

Medline entries

89098903:
Human interleukin 7: molecular cloning and growth factor
activity on human and murine B-lineage cells.

Peptide information for frame 2

ORF from 368 bp to 679 bp; peptide length: 104
Category: similarity to known protein

1 METSHAHESN CKIKGYGVVQ QLLHSQLCGA GEDLGPIGVS YVYSFRAVPF
51 SLILSNASLH SLGGKDSIQV GCLKELMGPL SELADQILGN LFNFYDFPSH
101 ILKW

BLASTP hits

Entry B32223 from database PIR:
interleukin-7 precursor (clone 1) - human
Score = 66, P = 7.0e-01, identities = 21/70, positives = 33/70

Alert BLASTP hits for DKFZphtes3_35e21, frame 2

PIR:B32223 interleukin-7 precursor (clone 1) - human, N = 1, Score =
66, P = 0.72

TREMBL:PADAL1_1 gene: "dall"; P.abies dall mRNA, N = 2, Score = 59, P
= 0.77

PIR:C32223 interleukin-7 precursor (clone 4) - human, N = 1, Score =
66, P = 0.79

TREMBL:PRU76726_1 gene: "PrMADS3"; product: "MADS-box protein"; Pinus
radiata MADS-box protein (PrMADS3) mRNA, complete cds., N = 2, Score =
59, P = 0.94

>PIR:B32223 interleukin-7 precursor (clone 1) - human
Length = 133

HSPs:

Score = 66 (9.9 bits), Expect = 1.3e+00, P = 7.2e-01
Identities = 21/68 (30%), Positives = 33/68 (48%)

Query: 39 VSYVYSFRAVPFSLIL-----SNASLHSLGGK--DSIQVGCLELMGPLSELADQILGNL 91
VS+ Y F P L+L S+ + GK +S+ + + +L+ + E+ L N
Sbjct: 4 VSFRIYFGLPPLILVLLPVASSDCDIEGKDGKQYESVLMVSIQQLLDSMKEIGSNCLNNE 63

Query: 92 FNFYDFPSHI 101
FNF F HI
Sbjct: 64 FNF--FKRHI 71

Pedant information for DKFZphtes3_35e21, frame 2

Report for DKFZphtes3_35e21.2

[LENGTH]	104
[MW]	11339.12
[pI]	5.87
[PROSITE]	MYRISTYL 2
[PROSITE]	PKC_PHOSPHO_SITE 1
[PROSITE]	ASN_GLYCOSYLATION 1
[KW]	Alpha_Beta

SEQ METSHAHESNCKIKGYGVVQQLLHSQLCGAGEDLGPIGVS YVYSFRAVPFSLILSNASLH
PRD ccchhhhhccccccchhhhhhhhhcc

SEQ SLGGKDSIQVGCLKELMGPLSELADQILGNLFNFYDFPSHILKW
 PRD cccccceeeccccccccccchhhhhhhhhccccccccccccccccc

Prosites for DKFZphtes3_35e21.2

PS00001	56->60	ASN_GLYCOSYLATION	PDOC00001
PS00005	44->47	PKC_PHOSPHO_SITE	PDOC00005
PS00008	63->69	MYRISTYL	PDOC00008
PS00008	89->95	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_35e21.2)

DKFZphtes3_35g6

group: testes derived

DKFZphtes3_35g6 encodes a novel 482 amino acid protein with high partial similarity to H. sapiens chromosome 19, cosmid R27216.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

strong similarity to R27216_1

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: /map="15"

Insert length: 3177 bp

Poly A stretch at pos. 3167, polyadenylation signal at pos. 3148

```

1 GGAGGCAGCG CCGGCCTCCG GAGGCGGCCT GGGCGATGGC GCGGAGTTT
51 TGTCCATAAC CTGGGCAACC GCGCAGCTGG AGGATGGCCT CACTCGGGCC
101 TGCCGCAGCT GGGGAGCAGG CGTCGGGGGC TGAGGCGGAG CCGGGCCCCG
151 CGGGGCCGCG GCCGCGCGCC TCACCGTCCT CTCTGGGGCC CTGCTCCCC
201 CTGCAGCGGG AACCTCTCTA CAACTGGCAG GCGACCAAGG CGTCGCTGAA
251 GGAGCGCTTC GCCTTCCTCT TCAACTCGGA GCTGCTGAGC GATGTGCGCT
301 TCGTACTGGG CAAGGCTCGC GCGCGCGCCG CCGCTGGGGG CCGCAGCGC
351 ATCCCCGCCC ACCGCTTCGT GCTGGCGGCC GGCAGCGCCG TCTTTGACGC
401 CATGTTCAAC GCGGCGCATG CCACCACGTC GGCCGAGATC GAGCTGCCGG
451 ACGTGGAGCC CGCAGCCTTC CTGGCGCTGC TGAGATTTCT ATATTGAGAT
501 GAAAGTTCAA TTGGTCCAGA AACAGTTATG ACCACTCTTT ATACTGCCAA
551 GAAATACGCA GTCCCAGCCT TGGAAGCACA CTGTGTAGAA TTCTCACCA
601 AACATCTTAG GGCAGATAAT GCCTTTATGT TACTTACTCA GGCTCGATTA
651 TTTGATGAAC CTCAGCTTGC TAGTCTTTGT CTAGATACAA TAGACAAAAG
701 CACAATGGAT GCAATAAGTG CAGAAGGGTT TACTGATATT GATATAGATA
751 CACTCTGTGC AGTTTTAGAG AGAGACACAC TCAGTATTCG AGAAAGTCGA
801 CTTTTTGGAG CTGTTGTACG CTGGGCAGAA GCAGATGTTC AGAGACAACA
851 ATTACCTGTG ACTTTTGGGA ATAAACAAAA AGTTCTAGGA AAAGCACTTT
901 CCTTAATCCG GTTCCCACTG ATGACAATTG AGGAATTTGC AGCAGGTCCCT
951 GCTCAATCTG GAATTTTGTG AGATCGTGAA GTGGTAAACC TCTTTCTTCA
1001 TTTTACTGTG AACCTTAAAC CCCGAGTTGA ATACATTGAC CGACCAAGAT
1051 GCTGTCTCAG GGGAAAGGAA TGCTGCATCA ATAGATTCCA GCAAGTAGAA
1101 AGCCGCTGGG GTTACAGTGG GACGAGTGAT CGAATCAGAT TCACAGTTAA
1151 TAGAAGGATC TCTATAGTTG GATTTGGCTT GTATGGATCT ATTCATGGCC
1201 CTACAGATTA TCAAGTGAAT ATACAGATCA TTGAATATGA GAAAAAGCAA
1251 ACCCTGGGAC AGAATGATAC CGGCTTTAGT TGTGATGGGA CAGCTAACAC
1301 ATTCAGGGTC ATGTTCAAGG AACCCATAGA GATCCTGCCC AATGTGTGCT
1351 ACACAGCATG TGCAACACTC AAAGGTCCAG ATTCCCACTA TGGCACAATA
1401 GGATTGAAGA AAGTAGTGCA TGAGACACCT GCTGCAAGCA AGACTGTTTT
1451 TTTCTTTTTT AGTTCCCTTG GCAATAATAA TGGCACTTCA ATAGAAGATG
1501 GACAAATCCG AGAAATCATA TTTTATACAT AATTTAGCAT TATAATACAT
1551 CTTGGCTAAA TAATACCATA CAATCTAGTG TCAAAAACAT AAATGGCCAC
1601 AAAAAAGTAG TTTGAGTGT ATGAATATTT AAAATTGTAA GATAAGAAAC
1651 AGTTTCTTAG AGCAGATAGA AAAATGCTTA TTTAAATCTT TGCATGATTT
1701 AAAAAACAGT TTTCCATTTT CTTACAACCT TAAGAGAAAA GAACTGGGTT
1751 TAATGGTTTA AAAAAAGCA CAGCTTTTTC ACCTTCATCT TGTATAATTT
1801 CATAGATTGG CTGACTTAGG GTCTTTCAAT AGTTTGGGAA TTGAAAGATT
1851 CTTGTTATAT ATAGCTAGTT TGGGTTTGT TTTGTTTTAA CTATTTTGAA
1901 GGTTAGGTGA GATGGGCAA TAGGCTTAAC TATTTGAAG GTTGGATGAA
1951 AAGAGATGGG TCAGTATTCC TACAGAATTC TTATTAACCT AAATAACTAA
2001 ATTTGAGAAA ATTAAGAAGC TGACTTTATA TTTGGTGGTT TGAAATATCT
2051 TGTTGTTAGC ATTTGTAATA ATGCTAAAAA AGGCCATAA AAATGCCCAA
2101 GAAAATATTC AGTGCAATTA TAGAGAAGGA TATTTGTAG TAGTATAGTA
2151 ATGTGTTATG TAGTACAGTT TTAAAGCTAT AAATGGAATT TTGTGTAAAT
2201 TCACAAAATG GTGATATAAA CAGGATCTAA GACTGGATTG CCTGTCACTA
2251 AACTGCACCA CTATACCTGT CTCTCTGTGT GGGGACACT GCTGATGATT
2301 CCCAAGATTG AGATGATGAC GGTGATGAC ACTGGTGAA CAGCCATCAC
2351 TTCAACATTG TGATAATCCT TCACAGCAAG AAACCGAATA AAATACTAAC
2401 ATTTCTAACA ACTGCTCTGA CATTGTAAAG AGATCCAACA GAATCACTCC
2451 TGCTGAAAAA TACGCTTTCT GCCACCTACA CATTCTATT TAGGAAGTAA
2501 AATTTGCTTC ATGGTCATGA CCCCATTAGT CAGTGTTACA GCTGTGTTGG
2551 GGATAGGAAG TATATCTGGC AGATTGACAT TTATACACT TTTTATAAAG
2601 CAGATTTTAA AATATAGTAA CATCCATTTT TTCCCTTGA AAGTGATTCT
2651 CTTATAAAAA ATGAAAGTGG AGTTTAAGGT ATATCAAATC GTTGTGGAAG
2701 GTGATTAAAA ATCAAAATTC TTTTAAATAT CAACTTAATT TTTTCTAAGT

```

```

2751 AAGATACAAA AAATTTTCAT CTAAAGTAAT ATTTCACTTT ATATTGTAAA
2801 GAAGGTAGGT ATATTGGTGG CTGAGGTCTC TTGAAATGTC TAAAGGGAAA
2851 TTTTCTATG GTAATGCTCT TACGGATATA AGCCTCAGTT AAATGGAATT
2901 ATCTATGGGA TGTGTGGTTC TGGTAACTA AAAATTAACC AGTAAACACT
2951 CTGTAGTAAC CATTACAGAA AATACTTCTG CCTTAAAAA TATGATATGC
3001 CAGAGATGAG TTAGTGTTTC TTGACGTGG AGACCTATAA ATGCCTCATC
3051 TGTGTACTG AACAAATTGAA ACTGCATGCA GCCATAAAG GGACAAGAAA
3101 CAGAATGTT TACTAATTT GGGACATCCC CTGGAGTTT TAAAAATAAA
3151 TAAATATATA TATATATAA AAAAAA

```

BLAST Results

Entry G37753 from database EMBL:
 SHGC-63477 Human Homo sapiens STS genomic.
 Score = 1627, P = 3.0e-66, identities = 327/329

Entry G37752 from database EMBL:
 SHGC-63476 Human Homo sapiens STS genomic.
 Score = 1578, P = 6.2e-64, identities = 320/324

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 84 bp to 1529 bp; peptide length: 482
 Category: similarity to unknown protein

```

1 MASLGPAAG EQASGAEEAP GPAGPPPPPS PSSLGPLLPL QREPLYNWQA
51 TKASLKERFA FLNSELLSD VRFVLGKGRG AAAAGGPORI PAHRFVLAAG
101 SAVFDAMFNG GMATTSAEIE LPDVEPA AFL ALLRFLYSDE VQIGPETVMT
151 TLYTAKKYAV PALEAHCVEF LTKHLRADNA FMLLTQARLF DEPLASLCL
201 DTIDKSTMDA ISAEGFTDID IDTLCAVLER DTLSIRESL FGAVVRWAEA
251 ECQRQQLPVT FGNKQKVLGK ALSLIRFLM TIEFAAGPA QSGILSDREV
301 VNLFLHFTVN PKPRVEYIDR PRCCLRGKEC CINRFQQVES RWGYSGTSDR
351 IRFTVNRRIIS IVGFGLYGSI HGPTDYQVNI QIIEYEKKQT LGQNDTGFSC
401 DGTANTFRVM FKEPIEILPN VCYTACATLK GPDSHYGTKG LKKVVHETPA
451 ASKTVFFFFS SPGNNNGTSI EDGQIPEIIF YT

```

BLASTP hits

Entry AC005306_2 from database TREMBL:
 product: "R27216.1"; Homo sapiens chromosome 19, cosmid R27216,
 complete sequence.
 Score = 1298, P = 1.9e-132, identities = 245/297, positives = 268/297

Entry CEF38H4.9 from database TREMBLNEW:
 gene: "F38H4.7"; Caenorhabditis elegans cosmid F38H4
 Score = 1237, P = 5.6e-126, identities = 248/446, positives = 322/446

Entry AC004678.1 from database TREMBL:
 product: "R34094.1"; Homo sapiens chromosome 19, cosmid R34094,
 complete sequence.
 Score = 555, P = 1.0e-53, identities = 112/137, positives = 123/137

Alert BLASTP hits for DKFZphtes3_35g6, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_35g6, frame 3

Report for DKFZphtes3_35g6.3

```

[LENGTH] 482
[MW] 52771.47
[pI] 5.79

```

[HOMOL] TREMBL:AC005306_2 product: "R27216_1"; Homo sapiens chromosome 19, cosmid R27216, complete sequence. 1e-142
 [BLOCKS] BL01075D Acetate and butyrate kinases family proteins
 [SUPFAM] POZ domain homology 3e-08
 [SUPFAM] A55R protein middle region homology 5e-06
 [SUPFAM] A55R protein 5e-06
 [SUPFAM] A55R protein carboxyl-terminal homology 5e-06
 [PROSITE] MYRISTYL 6
 [PROSITE] CAMP_PHOSPHO_SITE 2
 [PROSITE] CK2_PHOSPHO_SITE 9
 [PROSITE] TYR_PHOSPHO_SITE 1
 [PROSITE] PKC_PHOSPHO_SITE 7
 [PROSITE] ASN_GLYCOSYLATION 2
 [KW] Alpha_Beta
 [KW] LOW_COMPLEXITY 11.20 %

SEQ MASLGFAAAGEQASGAEEPGPAGPPPPSPSSLGPLLQREPLYNWQATKASLKERFA
 SEGXX.....
 PRD cccccccchhhhhhhccccccccccccccccccccccccccccchhhhhhhhhhhhh

SEQ FLFNSELLSDVRFVLGKGRGAAAAGGPQRIPAHRFVLAAGSAVFDAMFNGGMATTSAEIE
 SEGXXXXXXXXXXXX.....
 PRD hhhccccccceeeccccccccccccccccchhhheeeccccchhhhhhhhhcchhhhhhhhe

SEQ LPDVEPAFLALLRFLYSDEVQIGPETVMTTLTYAKKYAVPALEAHCVEFLTCKHLRADNA
 SEG
 PRD eccccchhhhhhhhhhhccccceechhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccch

SEQ FMLLTQARLFDEPQLASLCLDTIDKSTMDAISAEFTDIDITLCAVLERDTLSIRESRL
 SEG
 PRD hhhccchhhhhhhhhccccchhhhh

SEQ FGAVVRWAAEACQORQLPVTFGNKQKVLGKALSIRFPLMTIEEFAAGPAQSGILSDREV
 SEG
 PRD hhhhhhhhhhhhhhhhhhhhhccccchhh

SEQ VNLFLHFTVNPKEPRVEYIDRPRCCLRGKECCINRFQVESRWGSGTSDRIRFTVNRRIS
 SEG
 PRD hhhhheeeccccceeeccccceeeccccceehhhhhhhhhhhccccccccccccchhhhhhhceee

SEQ IVGFGLYGSIHGPTDYQVNIQIIEYEKKQTLGQNDTGFCSDGTANTFRVMFKEPIEILPN
 SEG
 PRD eeccccccccccccchhhhhhhcchhhhhhhhhccccccccccccccccceeeccccceeecccc

SEQ VCYTACATLKGPDSDHYGKGLKKVVHETPAASKTVFFFSSPGNNNGTSIEDGGQIPEIIF
 SEGXXXXXX.....
 PRD cceeeccccccccccccccccceeeccccceeeccccccccccccccccccccccccceec

SEQ YT
 SEG ..
 PRD CC

Prosite for DKFZphtes3_35g6.3

PS00001	394->398	ASN_GLYCOSYLATION	PDOC00001
PS00001	466->470	ASN_GLYCOSYLATION	PDOC00001
PS00004	357->361	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	387->391	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	54->57	PKC_PHOSPHO_SITE	PDOC00005
PS00005	154->157	PKC_PHOSPHO_SITE	PDOC00005
PS00005	234->237	PKC_PHOSPHO_SITE	PDOC00005
PS00005	296->299	PKC_PHOSPHO_SITE	PDOC00005
PS00005	348->351	PKC_PHOSPHO_SITE	PDOC00005
PS00005	406->409	PKC_PHOSPHO_SITE	PDOC00005
PS00005	428->431	PKC_PHOSPHO_SITE	PDOC00005
PS00006	14->18	CK2_PHOSPHO_SITE	PDOC00006
PS00006	54->58	CK2_PHOSPHO_SITE	PDOC00006
PS00006	115->119	CK2_PHOSPHO_SITE	PDOC00006
PS00006	206->210	CK2_PHOSPHO_SITE	PDOC00006
PS00006	217->221	CK2_PHOSPHO_SITE	PDOC00006
PS00006	234->238	CK2_PHOSPHO_SITE	PDOC00006
PS00006	281->285	CK2_PHOSPHO_SITE	PDOC00006
PS00006	296->300	CK2_PHOSPHO_SITE	PDOC00006
PS00006	468->472	CK2_PHOSPHO_SITE	PDOC00006
PS00007	430->437	TYR_PHOSPHO_SITE	PDOC00007
PS00008	80->86	MYRISTYL	PDOC00008
PS00008	110->116	MYRISTYL	PDOC00008
PS00008	365->371	MYRISTYL	PDOC00008

WO 01/12659

PCT/IB00/01496

PS00008	392->398	MYRISTYL	PDOC00008
PS00008	402->408	MYRISTYL	PDOC00008
PS00008	463->469	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_35g6.3)

DKFZphtes3_35k16

group: metabolism

DKFZphtes3_35k16 encodes a novel 666 amino acid protein with weak similarity to fatty acid-CoA synthetases/ligases.

The novel protein contains a putative AMP-binding domain signature, which is present in enzymes, which act via an ATP-dependent covalent binding of AMP to their substrate. This domain is found in several CoA synthetases, such as acetate-CoA ligase (EC 6.2.1.1), long-chain-fatty-acid-CoA ligase (EC 6.2.1.3), bile acid-CoA ligase. Therefore it is a new fatty acid-CoA synthetase/ligase with unknown substrate.

The new protein can find application in modulation of fatty acid metabolism and as a new enzyme for biotechnologic production processes.

similarity to acyl-CoA synthetase

complete cDNA, complete cds, potential start codon at Bp 50, few EST hits, seems to be a testis specific cDNA, 5 of 6 EST hits are from testis derived libraries

Sequenced by DKFZ

Locus: unknown

Insert length: 2520 bp

Poly A stretch at pos. 2510, polyadenylation signal at pos. 2490

```
1 CAGATGTCCC AGCTCCAGTG CTGTGGAGCA TGGTTTCTGC ACACCTGGAA
51 TGAAGTGAAC CCCAAGAGACT CAAGAAGGAG CTAAAGATCT TGAAGTAGAC
101 ATGAATAAAA CAGAAGTTAC TCCAGGCTG TGGACCACCT GTCGAGATGG
151 AGAAGTCCTT CTGAGGCTAT CCAACACGG ACCAGGCAT GAGACCCGA
201 TGACCATCCC TGAATTTTTT CGAGAGTCAG TCAACCGATT TGGAACTTAT
251 CCAGCCCTCG CATCCAAGAA TGGCAAAAAG TGGGAAATTC TGAATTTCAA
301 CCAGTACTAT GAGGCTTGTC GGAAGGCTGC AAAATCCTTG ATCAAGCTGG
351 GTTGGAGCG TTTCCACGGA GTTGGTATCC TGGGTTTAA CTCTGCAGAG
401 TGGTTTATCA CTGCTGTTGG TGCCATCCTA GCCGGGGGTC TTTGTGTTGG
451 TATTTATGCC ACCAACTCTG CCGAGGCTTG TCAATATGTC ATCACTCATG
501 CCAAGTGAA CATCTTGCTG GTTGAGATG ATCAACAGTT ACAGAAAATC
551 CTTTCGATT CACAGAGCAG CCTAGAGCCC CTAAAGCGA TCATCCAGTA
601 CAGACTGCCA ATGAAGAAGA ACAACAACCT GTACTCTTGG GATGATTTC
651 TGGAACTTGG CAGAAGTATC CCTGACACCC AACTGGAGCA GGTTCATCGAG
701 AGCCAGAAG CGAATCAATG CGCAGTGCTC ATCTACACTT CAGGGACCAC
751 AGGCATACCC AAGGGAGTGA TGCTCAGTCA TGACAACATC ACGTGGATTG
801 CAGGAGCAGT GACAAAGGAC TTTAAACTGA CAGACAAGCA TGAGACGGTG
851 GTTAGCTACC TCCCACTCAG CCATATTGCA GCACAGATGA TGGACATCTG
901 GGTACCCATA AAGATTGGGG CGCTCACATA CTTTGCTCAA GCAGATGCTC
951 TCAAGGGCAC CTTGGTAAAG ACTCTAAAGG AGGTAAACCT TACTGTCTTC
1001 ATTGGAGTGC CTCAAAATTG GGAGAAGATA CATGAGATGG TGAAGAAAAA
1051 TAGTGCCAAG TCCATGGGCT TGAAGAAGAA GGCATTCTGT TGGCAAGAA
1101 ACATTGGCTT CAAGGTCAAC TCAAAAAAGA TGTTGGGGAA ATATAATACT
1151 CCGGTGAGCT ACCGCATGGC TAAGACTCTC GTGTTCAAGC AAGTCAAGAC
1201 ATCCCTTGGC TTGGATCACT GTCACCTCTT TATCAGTGGG ACTCGCCCC
1251 TCAACCAAGA GACTGCCGAG TTCTTTCTAA GCTTGGACAT ACCTATAGGC
1301 GAGTTGTATG GGTGAGTGA GAGCTCGGGA CCCACACGA TATCCAACCA
1351 GAATAACTAC AGGCTTCTAA GCTGTGGCAA GATCTTGACT GGGTGTAAGA
1401 ATATGCTGTT CCAGCAGAAC AAGGATGGCA TTGGGGAGAT CTGCCTCTGG
1451 GGTAGGCACA TCTTCATGGG CTATCTGGAA AGTGAGACTG AAACACAGA
1501 GGCCATCGAT GATGAAGGCT GGCTACACTC TGGGGATCTG GGCCAGCTGG
1551 ACGGTCTGGG TTTCTCTAT GTCACCGGCC ACATCAAAGA AATCCTTATC
1601 ACTGCTGGTG GTGAAAATGT GCCCCCCATT CCTGTTGAGA CCTTGGTTAA
1651 GAAGAAGATC CCCATCATCA GTAACGCCAT GTTAGTAGGA GATAAACTGA
1701 AGTTTCTGAG CATGTTGCTG ACGCTGAAGT GTGAGATGAA TCAGATGAGC
1751 GGAGAACCCT TGGACAAGCT GAACCTCGAG GCCATCAACT TCTGTCGGGG
1801 TCTGGGCAGC CAGGCATCCA CCGTGACTGA GATGGTGAAG CAGCAAGACC
1851 CCTTGCTCTA CAAGGCCATC CAGCAAGGCA TCAATGCTGT GAACCAAGGA
1901 GCCATGAACA ATGCACAGAG GATTGAAAAG TGGGTCATCT TGGAGAAGGA
1951 CTTTTCCATC TATGGTGGAG AGCTAGGTCC AATGATGAAA CTTAAGAGAC
2001 ATTTTGTAGC CCAGAAATAC AAAAAACAAA TTGATCACAT GTACCACTGA
2051 CTGCTTTGAT GGAGCTGCTC TCAGCTGTTT TGATGCCTTC AGCAGGAAGA
2101 CCTCATTGCA ATAAGTGAAG TGCTGCTCTA GGTAGAAGCT CTCCTGCTG
2151 TTTTTAAGAA GCCACATTCC TCATTGGTCA GTTCTTGTAT TGTTCGCTG
2201 TTGGAGAGGT GCTCCCTAGA AGAACCTGCC ATACGTTTCA AAGCAATAAA
2251 ATCACTGTAT ATCTTTCTAA GGACCTTCAA GTCATGACTC CAGGGAAGCC
2301 TATTGGGAAG TCTACTAAAA ACTGCCTGAT TTACAAGAAA GACCTGAACT
```

```

2351 TGTGGGCTCC CATTGATT TTTTCTCCTC AGGGGACTCA GACATTAGAA
2401 AGAAAAAGCC TCACAGATT GAAGAACTGG ACCCCCAAT CAACTCACCT
2451 CCCTGGAAGC AACTGGGAAA CCCTTCCAAT AAGTCCTGAT AATAAAGCAC
2501 TTCAGGGTCC AAAAAAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 50 bp to 2047 bp; peptide length: 666
 Category: similarity to known protein

```

1 MTGTPKQEG AKDLEVDMNK TEVTPRLWTT CRDGEVLLRL SKHGPQHETP
51 MTIPEFFRES VNRFGTYPAL ASKNGKKWEI LNFNQYYEAC RKAASLIKL
101 GLERFHGVTG LGFNSAEWFI TAVGAILAGG LCVGIYATNS AEACQYVITH
151 AKVNILLVEN DQQLQKILSI PQSSLEPLKA IIQYRLPMKK NNNLYSWDDF
201 MELGRSIPDT QLEQVIESQK ANQCAVLIYT SGTGIPKGV MLSDNITWI
251 AGAVTKDFKL TDKHETVVSY LPLSHIAAQM MDIWPVIKIG ALTYFAQADA
301 LKGTILVSTLK EVKPTVFIGV PQIWEKIHEN VKKNSAKSMG LKKKAFVWAR
351 NIGFKVNSKK MLGKYNTGPS YRMAKTLVFS KVKTSGLLDH CHSFISGTAP
401 LNQETAFFFL SLDIPIGELY GLSESSGPHT ISNQNNYRL SCGKILTCK
451 NMLFQKNKDG IGEICLWGRH IFMGYLESET ETTEAIDDEG WLHSGDLGQL
501 DGLGFLYVTG HIKEILITAG GENVPPIPVE TLVKKKIPII SNAMLVGDKL
551 KFLSMLLTLL CEMNQMSGEP LDKLNFEAIN FCRGLGSQAS TVTEMVKQQD
601 PLVYKAIQGG INAVNQEAMN NAQRIEKWVI LEKDFSIIYG ELGPMMLKKR
651 HFVAQKYKKQ IDHMYH

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_35k16, frame 2

TREMBL:AB014531_1 gene: "KIAA0631"; product: "KIAA0631 protein"; Homo sapiens mRNA for KIAA0631 protein, partial cds., N = 1, Score = 1641, P = 8.9e-169

PIR:E70937 probable fadD15 - Mycobacterium tuberculosis (strain H37RV), N = 2, Score = 532, P = 3.6e-62

PIR:H64041 long-chain-fatty-acid--CoA ligase homolog - Haemophilus influenzae (strain Rd KW20), N = 2, Score = 486, P = 6.5e-59

>TREMBL:AB014531_1 gene: "KIAA0631"; product: "KIAA0631 protein"; Homo sapiens mRNA for KIAA0631 protein, partial cds.
 Length = 634

HSPs:

Score = 1641 (246.2 bits), Expect = 8.9e-169, P = 8.9e-169
 Identities = 319/628 (50%), Positives = 440/628 (70%)

```

Query: 38 LRLSKHGPQHETPMTIPEFFRESVNRFGTYPALASKNGKKWEILNFNQYYEACRKAASL 97
      LR+ P + P T+ F E+++++G AL K KWE ++++QYY R+AAK
Sbjct: 2 LRIDPSCP--QLPYTVHRMFYEALDKYGDLIALGFKRQDKWEHISYSQYYLLARRAAKGF 59

Query: 98 IKLGLERFHGVTGILGFSNAEWFITAVGAILAGGLCVGIYATNSAEACQYVITHAKVNILL 157
      +KLGL++ H V ILGFNS EWF +AVG + AGG+ GIY T+S EACQY+ N+++
Sbjct: 60 LKLGLQAHSAVAILGFNSPEWFFSAVGTVFAGGIYTTSSPEACQYIAYDCCANVIM 119

Query: 158 VENDQQLQKILSIPOSSLEPLKAIQYRLPM-KKNNNLYSWDDFMELGRSIPDTQLEQVI 216
      V+ +QL+KIL I L LKA++ Y+ P K N+Y+ ++FMELG +P+ L+ +I
Sbjct: 120 VDTQKQLEKILKI-WKQLPHLKAVVIYKEPPPNKMANVYTMEEFMELGNEVPPEALDAII 178

Query: 217 ESQKANQCAVLIYTS GTTGIPKGVMLSDNITWIA--GAVTKDFKLTD-KHETVVSYLPL 273

```

```

++Q+ NQC VL+YTS GTTG PKGVMLS DNITW A G+ D + + + E VVSYLPL
Sbjct: 179 DTQQPNQCCVLVYTS GTTGPNKGVMLSQDNITWTARYGSQAGDIRPAEVQQEVVSYLPL 238

Query: 274 SHIAAQMMDIWVPIKIGALTYFAQADALKGTLVSTLKEVKPTVFIGVPQIWEKIHVMVK 333
SHIAAQ+ D+W I+ GA FA+ DALKG+LV+TL+EV+PT +GVP++WEKI E +++
Sbjct: 239 SHIAAQIYDLWTGIQWGAQVCFAEPDALKGS LVNTLREVEPTSHMGVPRVWEKIMERIQE 298

Query: 334 NSAKSMGLKKKAFVWARNIGFKVNSKKMLGKYNTPVSYRMAKTLVFSKVKTSGLGDHCHS 393
+A+S +++K +WA ++ + N G P + R+A LV +KV+ +LG C
Sbjct: 299 VAAQSGFIRRKMLLWAMSVTLEQNLT-CPGSDLKPF TTRLADYLV LAKVRQALGFACQK 357

Query: 394 FISGTAPLNQETAFFLSLDIPIGELYGLSESSGPHTISNQNNYRLSCGKILTGCKNML 453
G AP+ ET FFL L+I + YGLSE+SGPH +S+ NYRL S GK++ GC+ L
Sbjct: 358 NFYGAAPMMAETQHFFLGLNIRLYAGYGLSETSGPHFMSSPYNYRLYSSGKLVPGCRVKL 417

Query: 454 FQONKDGIGEICLWGRHIFMGYLESETETTEAIDDEGLHSGDLGLDGLGFLYVTGHIK 513
Q+ +GIGEICLWGR IFMGYL E +T EAID+EGWLH+GD G+LD GFLY+TG +K
Sbjct: 418 VNQDAEGIGEICLWGRITIFMGYLNMEDKTCEAIDEEGLHTGDAGRLDADGFLYITGRK 477

Query: 514 EILITAGGENVPPPIPVETLVKKKIPISNAMLVGDKLFLSMLLTLCENMQMSGEPLDK 573
E++ITAGGENVPP+P+E VK ++PIISNAML+GD+ KFLSMLLTLC ++ + + D
Sbjct: 478 ELIITAGGENVPPVPIEEAVKMELPISNAMLIGDQRKFLSMLLTLCCTLDPDTS DQTDN 537

Query: 574 LNFEAINFCRGLGSQASTVTEMVKQDPLVYKAIQQGINAVNQEAMNNAQRIEKWVILEK 633
L +A+ FC+ +GS+A+TV+E++++D VY+AI++GI VN A I+KW ILE+
Sbjct: 538 LTEQAVEFCQRVGS RATTVSEIIEKKDEAVYQAIIEGIRRVNMNAAARPYHIQKWAILER 597

Query: 634 DFSIYGELGPMMLKLRHFVAQKYKQIDHMY 665
DFSI GGELGP MKLKR V +KYK ID Y
Sbjct: 598 DFSISGELGPTMKLR LTVLEKYKGIIDSFY 629

```

Pedant information for DKFZphtes3_35k16, frame 2

Report for DKFZphtes3_35k16.2

```

[LENGTH]      666
[MW]           74344.97
[pI]           8.67
[HOMOL]        TREMBL:AB014531_1 gene: "KIAA0631"; product: "KIAA0631 protein"; Homo sapiens
mRNA for KIAA0631 protein, partial cds. 1e-176
[FUNCAT]       i lipid metabolism [H. influenzae, HI0002] 2e-55
[FUNCAT]       08.10 peroxisomal transport [S. cerevisiae, YER015w] 2e-29
[FUNCAT]       30.19 peroxisomal organization [S. cerevisiae, YER015w] 2e-29
[FUNCAT]       01.06.13 lipid and fatty-acid transport [S. cerevisiae, YER015w] 2e-29
[FUNCAT]       01.06.07 lipid, fatty-acid and sterol utilization [S. cerevisiae, YER015w]
2e-29
[FUNCAT]       01.06.01 lipid, fatty-acid and sterol biosynthesis [S. cerevisiae, YMR246w]
2e-23
[FUNCAT]       06.07 protein modification (glycosylation, acylation, myristylation,
palmitoylation, farnesylation and processing) [S. cerevisiae, YMR246w] 2e-23
[BLOCKS]       BL00455
[SCOP]         dl1ci 5.19.1.1.1 Luciferase [Firefly (Photinus pyralis) 1e-49
[EC]           1.13.12.7 Photinus-luciferin 4-monooxygenase (ATP-hydrolysing) 9e-17
[EC]           6.2.1.3 Long-chain-fatty-acid--CoA ligase 4e-34
[EC]           5.1.1.11 Phenylalanine racemase (ATP-hydrolysing) 6e-08
[EC]           6.2.1.12 4-Coumarate--CoA ligase 8e-18
[PIRKW]        duplication 6e-07
[PIRKW]        phosphopantetheine 3e-12
[PIRKW]        multifunctional enzyme 3e-06
[PIRKW]        ligase 6e-08
[PIRKW]        acid-thiol ligase 4e-34
[PIRKW]        transmembrane protein 5e-22
[PIRKW]        monooxygenase 9e-17
[PIRKW]        hydrolase 4e-34
[PIRKW]        peroxisome 9e-15
[PIRKW]        antibiotic biosynthesis 3e-12
[PIRKW]        isomerase 6e-08
[PIRKW]        flavonoid biosynthesis 1e-17
[PIRKW]        magnesium 9e-15
[PIRKW]        ATP 5e-22
[PIRKW]        oxidoreductase 9e-17
[PIRKW]        liver 2e-31
[SUPFAM]        alpha-aminoacidipyl-cysteiny-valine synthetase 3e-07
[SUPFAM]        human long-chain-fatty-acid--CoA ligase 4e-34
[SUPFAM]        gramicidin S synthetase I 6e-08
[SUPFAM]        peptide synthetase ppsE 7e-06
[SUPFAM]        gramicidin S synthetase I repeat homology 3e-12
[SUPFAM]        peptide synthetase ppsD 2e-07

```

```

[SUPFAM]    probable acyl-CoA ligase medium chain 2e-09
[SUPFAM]    acetate--CoA ligase 8e-10
[SUPFAM]    acetate--CoA ligase homology 4e-54
[SUPFAM]    surfactin synthetase 3e-12
[SUPFAM]    4-coumarate--CoA ligase 8e-18
[SUPFAM]    short-chain alcohol dehydrogenase homology 8e-07
[SUPFAM]    acyl carrier protein homology 2e-29
[PROSITE]   MYRISTYL      12
[PROSITE]   AMP_BINDING   1
[PROSITE]   AMIDATION     1
[PROSITE]   CAMP_PHOSPHO_SITE 1
[PROSITE]   CK2_PHOSPHO_SITE 9
[PROSITE]   TYR_PHOSPHO_SITE 3
[PROSITE]   PKC_PHOSPHO_SITE 10
[PROSITE]   ASN_GLYCOSYLATION 2
[PFAM]      AMP-binding enzymes
[KW]        Irregular
[KW]        3D
[KW]        LOW_COMPLEXITY      1.80 %

SEQ  MTGTPKTOEGAKDLEVDNMNKTETVTPRLWTTTCRDGEVLLRLSKHGPGHETPMTIPEFFRES
SEG  .....
1lci- .....

SEQ  VNRFGTYPALASKNGKKWEILNFNQYYEACRKAASLIKGLERFHGVGILGFNSAEWFI
SEG  .....
1lci- .....

SEQ  TAVGAILAGGLCVGIYATNSAEACQYVITHAKVNILLVENDQQLQKILSIPQSSLEPLKA
SEG  .....
1lci- .....

SEQ  IIQYRLPMKNNNLYSWDDFMELGRSIPDTQLEQVIESQKANQCAVLIYTSGTTGIPKGV
SEG  .....
1lci- .....

SEQ  MLSHDNITWIAGAVTKDFKLTDKHETVVSYLPLSHIAAQMMDIWVPIKIGALTYFAQADA
SEG  .....
1lci- .....

SEQ  LKGTLVSTLKEVKPTVFIGVPQIWEKIHVMVKKNSAKSMGLKKKAFVWARNIGFKVNSKK
SEG  .....
1lci- .....

SEQ  MLGKYNTPVSYRMAKTLVFSKVKTSGLDCHSFISGTAPLNQETAEFFFLSLDIPIGELY
SEG  .....
1lci- .....TTTTCEEETTTTCCCHHHHHHHHHHCCCCBCEE

SEQ  GLSESSGPHITISNQNNYRLSCGKILTGCKNMLFQQNKDGIGEICLWGRHIFMGYLESET
SEG  .....
1lci- ECGGGTTEEEECCEEEEEETTTTEEEETTTTCEEETTEEEETTTTCEEETTHH

SEQ  ETTEAIDDEGLWLSGDLGLDGLFLYVTGHIKEILITAGGENVPPIPVETLVKKKIPII
SEG  .....XXXXXXXXXXXXX.....
1lci- HHHHHBTTTTCEEEEEEETTTTCEE-----ECEETTEEECHHHHHHHHHHT-TTE

SEQ  SNAMLVGDKLKFSLMLTLKCEMNQMSGEPLDKLNFEAINFCRGLGSQASTVTEMVKQQD
SEG  .....
1lci- EEEEEEE.....

SEQ  PLVYKAIQQGINAVNQEAMNNAQRIEKWVILEKDFSIIYGELGPMMLKRRHFVAQYKKQ
SEG  .....
1lci- .....

SEQ  IDHMYH
SEG  .....
1lci- .....

```

Prosites for DKFZphtes3_35k16.2

PS00001	19->23	ASN_GLYCOSYLATION	PDOC00001
PS00001	246->250	ASN_GLYCOSYLATION	PDOC00001
PS00004	332->336	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	4->7	PKC_PHOSPHO_SITE	PDOC00005
PS00005	24->27	PKC_PHOSPHO_SITE	PDOC00005
PS00005	30->33	PKC_PHOSPHO_SITE	PDOC00005
PS00005	218->221	PKC_PHOSPHO_SITE	PDOC00005
PS00005	261->264	PKC_PHOSPHO_SITE	PDOC00005

PS00005	308->311	PKC_PHOSPHO_SITE	PDOC00005
PS00005	335->338	PKC_PHOSPHO_SITE	PDOC00005
PS00005	358->361	PKC_PHOSPHO_SITE	PDOC00005
PS00005	370->373	PKC_PHOSPHO_SITE	PDOC00005
PS00005	558->561	PKC_PHOSPHO_SITE	PDOC00005
PS00006	30->34	CK2_PHOSPHO_SITE	PDOC00006
PS00006	52->56	CK2_PHOSPHO_SITE	PDOC00006
PS00006	173->177	CK2_PHOSPHO_SITE	PDOC00006
PS00006	196->200	CK2_PHOSPHO_SITE	PDOC00006
PS00006	206->210	CK2_PHOSPHO_SITE	PDOC00006
PS00006	210->214	CK2_PHOSPHO_SITE	PDOC00006
PS00006	308->312	CK2_PHOSPHO_SITE	PDOC00006
PS00006	478->482	CK2_PHOSPHO_SITE	PDOC00006
PS00006	591->595	CK2_PHOSPHO_SITE	PDOC00006
PS00007	659->666	TYR_PHOSPHO_SITE	PDOC00007
PS00007	658->666	TYR_PHOSPHO_SITE	PDOC00007
PS00007	597->605	TYR_PHOSPHO_SITE	PDOC00007
PS00008	3->9	MYRISTYL	PDOC00008
PS00008	65->71	MYRISTYL	PDOC00008
PS00008	124->130	MYRISTYL	PDOC00008
PS00008	130->136	MYRISTYL	PDOC00008
PS00008	134->140	MYRISTYL	PDOC00008
PS00008	235->241	MYRISTYL	PDOC00008
PS00008	239->245	MYRISTYL	PDOC00008
PS00008	303->309	MYRISTYL	PDOC00008
PS00008	387->393	MYRISTYL	PDOC00008
PS00008	421->427	MYRISTYL	PDOC00008
PS00008	498->504	MYRISTYL	PDOC00008
PS00008	586->592	MYRISTYL	PDOC00008
PS00009	74->78	AMIDATION	PDOC00009
PS00455	227->239	AMP_BINDING	PDOC00427

Pfam for DKFZphtes3_35kl6.2

HMM_NAME	AMP-binding enzymes
HMM	*TYRELNERANRLARHLRsekGlrPGDiVgIMMDRSMWMIVaMLGIWKAG + + +E +A L+ +G VGI+ +S + ++ G + AG
Query	82 NFNQYYEACRKAASLI-KLGLERFHVGILGFNSAEWFTAVGAILAG 129
HMM	GAYVPIDPeYPdERiQYMLEDSGARLLITQrh...HmqRIPdemwwvdH G +V I +E QY++ ++ + +L+++ + + IP++++ +
Query	130 GLCVGIYATNSAEACQYVITHAKVNILLVENDQQLQKILSIPQSSLEPLK 179
HMM	IivdWe.....WddlWWHedeeNppqWvdPeDLAYIIY +I++ + + +++++ + E ++ +++++ A +IY
Query	180 AIQYRLPMKNNNLYSWDDFMELGRSIPDTQLEQVIESQKANQCAVLIY 229
HMM	TSGTTGPKGVMIEHrNiVnycqWMnWRYGmteeDDRILWftSDpYWFDa TSGTTG PKGVM++H NI+ + + + +T+ + + + + + A
Query	230 TSGTTGIPKGVMLSHDNITWIAGAVTKDFKLTKHETVVSYP-LSHIAA 278
HMM	SVWDMFWpLLnGatLYIpPeEtRrDPerWWqYIqRHgITWWylTPSMFRM +++D++ P+ GA Y + ++ + +++++ ++T+ ++P +++
Query	279 QMMDIWVPIKIGALTYFAQADAL--KGTLVSTLKEVKPTVFIGVPQIWEK 326
HMM	Lmpd..... + +
Query	327 IHEMVKKNSAKSMGLKKKAFVWARNIGFKVNSKKMLGKYNTPVSYRMAKT 376
HMMpsLRhVMFgGEpLsPehWdWWRkrfgkgRIINMYWPT ++ + +++G PL++E+++ ++ + ++I Y+ +
Query	377 LVFSKVKTSGLDHCHSFISGTAPLNQETAEFFL-SLD--IPIGELYGLS 423
HMM	ETTVWtTwMrIiPdepeqWrwiPIGRPIpNTqWYIMDdnMQLPiGViGE E++ T+ + + R +++G+ + + + + +N G IGE
Query	424 ESSGPHTISNQNN--Y---RLSCGKILTGCNMLFQQN---KDG-IGE 463
HMM	LYigGWPGVARGYWNRPTELTEERfipNPFPGEYRrGWNrRMYRTGDLAR +++ G ++ GY+ + +T E+ + ++ ++GDL++
Query	464 ICLWG-RHIFMGYLESETETTEAIDDEGW-----LHSGDLGQ 499
HMM	WlPDGnIEYLGRIID.QVKIRGYRIELGEIEhqlr.qHPgIqEAVV* + G+++ G I + G+++ + +E+ + ++P I+ A
Query	500 LDGLGFLYVTGHIKEILITAGGENVPPIPVETLVKKKIPIISNAML 545

DKFZphtes3_35k24

group: transmembrane protein

DKFZphtes3_35k24 encodes a novel 514 amino acid protein without similarity to known proteins.

The novel protein contains 5 transmembrane regions.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes and as a new marker for testicular cells.

unknown ;

membrane regions: 5

Summary DKFZphtes3_35k24 encodes a novel 514 amino acid protein.

No homologues found in bacteria yeast and C.elegans, specific for mammals?

unknown

complete cDNA, complete cds, few EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 2706 bp

Poly A stretch at pos. 2696, polyadenylation signal at pos. 2675

```
1 CCGTGTGCAG TCGCCCCGCG CCCC CGCGCA CCCTTCGGGT AAAC TACGAA
51 CTGGGAGTTC TGAAGAATGG GTAAAGACTT TCGTTACTAT TTCCAGCATC
101 CCTGGTCTCG CATGATTGTG GCTTACTTGG TGATCTTCTT TAACTTCTTA
151 ATATTTGCGG AGGACCCAGT TTCTCATAGC CAAACAGAAG CCAATGTTAT
201 TGTGTGTTGA AACTGTTTTT CATTTGTTAC AAATAAATAC CCTAGAGGAG
251 TTGGCTGGAG GATTTTGAAG GTGCTTCTAT GGCTACTTGC CATTTCTACA
301 GGACTAATAG CTGGCAAATT TCTGTTCCAT CAGCGTTTGT TTGGTCAGTT
351 GCTCCGATTA AAAATGTTTC GAGAAGATCA TGGGTCGTGG ATGACAATGT
401 TCTTCAGCAC AATTCTCTTT CTCTTCATAT TTTCTCACAT ATACAACACG
451 ATTCTTCTAA TGGATGGGAA CATGGGAGCA TATATCATT A CAGACTATAT
501 GGGCATCCGA AATGAAAGTT TCATGAAATT AGCTGCAGTA GGGACCTGGA
551 TGGGGGACTT TGTCACAGCT TGGATGGTCA CTGATATGAT GCTTCAGGAC
601 AAACCCCTATC CTGACTGGGG AAAATCAGCA AGAGCTTCTT GGAAGAAAGG
651 AAATGTTAGG ATCACTTTAT TCTGGACAGT TCTTTTACT CTGACCTCTG
701 TGTTGTGATC TGTGATTACA ACGGACTGGA TCAGCTGGGA CAAGCTGAAT
751 CGGGGATTTT TGCCAGTGA TGAAGTTTCC AGAGCATTCC TTGCTTCTTT
801 TATCTTGGTC TTTGACCTTC TTATTGTGAT GCAGGACTGG GAATTCCTAC
851 ATTTTCATGG AGATGTTGAT GTAAATCTCC CTGGTTTGCA CACCCCTCAC
901 ATGCAGTTCA AGATTCTTTT CTTCAGAAA ATCTTCAAGG AGGAATATCG
951 TATTCACATA ACAGGCAAAT GGTTTAACTA TGGAAATATC TTCTCTGTCT
1001 TGATTTTGGG TCTTAATATG TGAAGAAGCC AAATATTTTA TAAACCTCAT
1051 GAATATGGGC AATATATCGG CCCGGGGCAG AAGATATATA CAGTGAAAGA
1101 CTCAGAAAGT TTAAGAGATT TGAACAGAAC CAAGCTATCC TGGGAATGGA
1151 GGTCCAATCA CACTAACCCT CGGACTAATA AAACATATGT TGAGGGAGAC
1201 ATGTTCTTAC ACAGCAGGTT CATAGGAGCC AGTCTTGATG TCAAGTGTCT
1251 GGCCTTTGTT CCAAGCCTGA TAGCCTTTGT GTGGTTTGGG TTCTTTATTT
1301 GGTCTTTTGG ACGATTTTGG AAAAAATGAGC CACGCATGGA GAATCAAGAC
1351 AAAACTTACA CTCGCATGAA AAGAAATCTT CCATCAGAAC ATAGCAAAGA
1401 CATGGGAATC ACTCGAGAAA ACACCCAGGC TTCAGTAGAA GACCCCTTGA
1451 ATGACCCCTC TTTGGTTTGC ATCAGGTCTG ACTTCAATGA GATCGTCTAC
1501 AAGTCTTCCC ACCTAACCTC GGAAAACCTG AGCTCACAGT TGAACGAATC
1551 TACTAGTGCA ACAGAAGCTG ATCAAGACCC AACGACTTCT AAAAGTACAC
1601 CTACGAACTA GACTCGGAGA TAGACTTGGG GATAACACAA AAAGCAACCT
1651 TGAGTGTAAC TTTAAAAAAT TAGTCTTTCC TTTTGTATAT GTAAGGTTTA
1701 CGTAGTGTTA GGTAAAAATA TGAACAATGC CACAACGGTG CTCAACATGC
1751 TTTTCTAGG ATTCAATTGT TTCTATTTGT ATTATAATAC ACGTGCCTAC
1801 TGTATACTCA ACAGTCTCTT AGAGATTGCT TTTCACAATT GCACAAGCTA
1851 TTTACTGACT TACAGCATAG TGAAGATTA GCTGATGACC CATGTATCTG
1901 ATGTTCAACC ATAGTGGTGC CTTGAGACAT TAACTGTTT TTAAGTGTAC
1951 CAGAAATGAA GTGTGGAAAC GTTACCTAAC CTATTTTACA TGGGCGTTTT
2001 GTATACAACT ATTTTGATCT ACACCTGATG TCTGAGCAGA AAACAGAAAT
2051 AGCTAAATGT GACTCAGGAA GTATCTCTTG GTTCTTATT CAGCAGCAGA
2101 GTTGGTGACT TTGACAACTG GACTGCAGAG AAACATGGTG ATCACCTTTT
2151 AATTTTTTAT GGCTGTCTGC CAAATATAAA TACAGATGCA AAATTCAGTA
2201 ATAGGAGATC CATAACCCAA CATGGGTAC TACTCGTGAA ATGTGACTTT
2251 CTCCCACCAG TAATTGAAAT GAGGTGATGA TACCTAATTA TGTTTTCTTA
2301 ATTAAGATA AATTGCTACT TGATTAAAAA TCCTGCCCTT CACCTTTGGG
```

2351	AACAAAGGTT	AAGAGACACA	GTTGGGCGAA	CTCTCAAATT	TATTGGCATT
2401	TACACAAAGT	CCGACCAACG	CAGGAACACT	AAGTTTTCAT	CATATGAGAG
2451	CAGCACAATC	CACCATTATC	AATATTCTGA	TATCTTTCGT	CAAAATATGC
2501	TCTGGATAGT	GAAAATTGAA	AAACATATGC	CAACCCCTGAG	CAGGGAACT
2551	CCATCAAAAA	TCATCGAGCG	GAACTTTGCT	AGGTAGAGAA	CGCTGCATG
2601	AAGAATAATT	TTTAATGCTT	TGTTTTGCGT	ATGTGTTTTT	TGTTTTTGTT
2651	TTTAAAGAAC	TAAATATTGC	ACATTAATAA	ATAAGAATTA	TACAGCAAAA
2701	AAAAAA				

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 67 bp to 1608 bp; peptide length: 514
Category: putative protein

1	MGKDFRYFYQ	HPWSRMIVAY	LVIFFNFLIF	AEDPVSHSQT	EANVIVVGN
51	FSEVTHKSWT	GVGWRILKVL	LWLLAILTGL	IAGKFLFHQR	LFQGLLRLLK
101	FREDHGNYMR	MFSTLTLKFL	IFSHYITLGL	LDMDGMGYAT	ITDYMGRIRN
151	SFMKLAUVGT	WMGDVFYAWM	VTDMLLQDKP	YPDWKGSAFA	FWKKGNVIRI
201	LFWTVLTFLT	SVFVVLVITD	WISWDKLNRR	FLPSPDEVSRA	FLASFLIVFD
251	LLIVMDQWEF	PHFMGVDDVN	PLGLHTPHMQ	KIPFFQKRI	KEEYRIHITG
301	KWENYGIJLF	VLLLDLNMPT	NOIFYQPHYE	GQYIPGGQKI	YTVKDESFLK
351	LDNRKTLGWE	WRSNHTNWRK	NKTYVEGDMF	LHSRFIGASL	YDVCKSLAFV
401	LIAFVWFGEF	IWFFGRFLKN	EPRMENQDKT	YTRMKRKSPS	EHSKDMGITR
451	ENTQASVSDS	LDNPSLVCI	SDFNEIVYKS	SHLTSENLS	QLNESTSATE
501	ADODPTTSK	TPTN			

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3 35k24, frame 1

No Alert BLASTP hits found

Pedant information for DKF2phtes3_35k24, frame 1

Report for DKFZphtes3_35k24.1

```
[LENGTH]          514
[MW]               60185.03
[pI]               8.67
[PROSITE]          MYRISTYL           5
[PROSITE]          CAMP_PHOSPHO_SITE      1
[PROSITE]          CK2_PHOSPHO_SITE       8
[PROSITE]          TYR_PHOSPHO_SITE       1
[PROSITE]          PKC_PHOSPHO_SITE       7
[PROSITE]          ASN_GLYCOSYLATION      6
[KW]               SIGNAL_PEPTIDE 32
[KW]               TRANSMEMBRANE  5
[KW]               LOW_COMPLEXITY      15.37 %
```

```

SEQ      MGKDFRYFYQHPWSRMIVAYLVIFNFNLI FAEDPVSHSQTEANVIVVGNCFSVTVNKYPR
SEG      .....
PRD      cccceeeeeecchhhhhhhhhhhhhhhhhhhccccccccccceeeeeecccccecccc
MEM      .....

SEQ      GVGWRILKVLLWLLAILTGLIAGKFLFHQRLFGQLRLKMFREDHGSWMTMFFSTILFLF
SEG      .....XXXXXXXXXXXXXXXXXXXXX.....XXXXXXXXXXXXXXXX
PRD      cchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccceeeehhhhhhhhh
MEM      MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

```



```

SEQ      IFSHIYNTILLMDGNMGAYIITDYMGI RNESFMKLA AVGTWMGDFVTAWMVTDMMLQDKP
SEQ      xxx.....
PRD      hhhhhhhhhhhcccccceeeccccchhhhhhhhhcccccchhhhhhhhhccccc
MEM      MMMMMMMMMMMM.....

SEQ      YPDWGKSARAFWKGNVRITLFWTVLFTLTSVVVLVITD WISWDLN RGF LPSDEVSA
SEQ      .....xxxxxxxxxxxxxxxxxxxxx.....
PRD      cccccchhhhhhhcccccceehhhhhhhhhheeeccccccccccccccccchhhh
MEM      .....MMMMMMMMMMMMMMMMMM.....M

SEQ      FLASFILVFDLLIVMQDWEFFPHMGD VDNLPGLLTHPMQFKIPFFQKIFKEEYRIHTG
SEQ      .....xxxxxxxxxxxxxxxxxxxxx.....
PRD      hhhhhhhhhhhhhhhhhcccccceccccccccccccccccccccchhhhhhhhhhhhhccccc
MEM      MMMMMMMMMMMMMMMMM.....

SEQ      KWFNYGII FLVLILD LNMWKNQIFYKPHEYGYIGPGQKIYTKDSESLKDLNRTKLSWE
SEQ      .....
PRD      cceeeeee hhhhhhhcccccceccccccccccccccccceeeccccccccccccchhhh
MEM      .....

SEQ      WRSNHTNPRTNKTYVEGDMFLHSRFIGASLDVKCLAFVPSLIAFWVFGFFIWFGRFLKN
SEQ      .....xxxxxxxxxxxxxxxxxxxxx.....
PRD      hhccccccccccccccccchhhhhcccccceeeehhhheeeccccceeeeeecccc
MEM      .....MMMMMMMMMMMMMMMMMM.....

SEQ      EPRMENQDKTYTRMKRKS PSEHSKDMGITRENTQASVEDPLNDPSLVCIRSDNFNIVYKS
SEQ      .....
PRD      cccccccccchhhhhcccccceeeccccccccccccccccceeeccccceeecc
MEM      .....

SEQ      SHLTSENLS SQLNESTSATEADQDPTTSKSTPTN
SEQ      .....
PRD      cccccccccccccccccccccccccccccccccccc
MEM      .....

```

Prosites for DKFZphtes3_35k24.1

PS000001	149->153	ASN_GLYCOSYLATION	PDOC000001
PS000001	353->357	ASN_GLYCOSYLATION	PDOC000001
PS000001	364->368	ASN_GLYCOSYLATION	PDOC000001
PS000001	371->375	ASN_GLYCOSYLATION	PDOC000001
PS000001	487->491	ASN_GLYCOSYLATION	PDOC000000
PS000001	493->497	ASN_GLYCOSYLATION	PDOC000001
PS000004	435->439	CAMP_PHOSPHO_SITE	PDOC000004
PS000005	55->58	PKC_PHOSPHO_SITE	PDOC000005
PS000005	187->190	PKC_PHOSPHO_SITE	PDOC000005
PS000005	299->302	PKC_PHOSPHO_SITE	PDOC000005
PS000005	342->345	PKC_PHOSPHO_SITE	PDOC000005
PS000005	348->351	PKC_PHOSPHO_SITE	PDOC000005
PS000005	370->373	PKC_PHOSPHO_SITE	PDOC000005
PS000005	507->510	PKC_PHOSPHO_SITE	PDOC000005
PS000006	38->42	CK2_PHOSPHO_SITE	PDOC000006
PS000006	342->346	CK2_PHOSPHO_SITE	PDOC000006
PS000006	348->352	CK2_PHOSPHO_SITE	PDOC000006
PS000006	373->377	CK2_PHOSPHO_SITE	PDOC000006
PS000006	438->442	CK2_PHOSPHO_SITE	PDOC000006
PS000006	456->460	CK2_PHOSPHO_SITE	PDOC000006
PS000006	497->501	CK2_PHOSPHO_SITE	PDOC000006
PS000006	499->503	CK2_PHOSPHO_SITE	PDOC000006
PS000007	326->334	TYR_PHOSPHO_SITE	PDOC000007
PS000008	48->54	MYRISTYL	PDOC000008
PS000008	79->85	MYRISTYL	PDOC000008
PS000008	106->112	MYRISTYL	PDOC000008
PS000008	134->140	MYRISTYL	PDOC000008
PS000008	159->165	MYRISTYL	PDOC000008

(No Pfam data available for DKFZphtes3 35k24.1)

DKFZphtes3_35n12

group: metabolism

DKFZphtes3_35n12 encodes a novel 315 amino acid protein with strong similarity to ADP,ATP carrier T (ANT) proteins.

The novel protein contains three mitochondrial energy transfer signatures and is closely related to the ADP/ATP translocator, or adenine nucleotide translocator (ANT), a protein most abundant in mitochondria. In its functional state, it is a homodimer of 30-kD subunits embedded asymmetrically in the inner mitochondrial membrane. The dimer forms a gated pore through which ADP is moved from the matrix into the cytoplasm.

The new protein can find application in modulation of ADP-transport and energy metabolism in cells/mitochondria.

strong similarity to ADP/ATP carrier proteins

EST hits to mouse and drosophila

Sequenced by DKFZ

Locus: unknown

Insert length: 1803 bp

Poly A stretch at pos. 1793, polyadenylation signal at pos. 1772

```
1 AGCGTCCCAA GAGCCACTTT CTCGCCAGTA CGATGCTGCA GCGGTTTTCC
51 GGTTTTCCGC TTCCCTTCAT CGTAGCTCCC GTACTCATTT TTAGCCACTG
101 CTGCCGGTTT TTATATCCTT CTCCATCATG CATCGTGAGC CTGCGAAAAA
151 GAAGGCAGAA AAGCGGCTGT TTGACGCCTC ATCCTTCGGG AAGGACCTTC
201 TGGCCGGCGG AGTCGCGGCA GCTGTGTCCA AGACAGCGGT GGCGCCCATC
251 GAGCGGGTGA AGCTGCTGCT GCAGGTGCAG GCGTCGTCGA AGCAGATCAG
301 CCCCAGAGCG CGGTACAAAG GCATGGTGGG CTGCCTGGTG CGGATTCCTC
351 GCGAGCAGGG TTTCTTCAGT TTTTGGCGTG GCAATTTGGC AAATGTTATT
401 CGGTATTTTC CAACACAAGC TCTAACTTTT GCTTTTAAGG ACAAATACAA
451 GCAGCTATTTC ATGTCTGGAG TTAATAAAGA AAAACAGTTC TGGAGGTGGT
501 TTTTGGCAAA CCTGGCTTCT GGTGGAGCTG CTGGGGCAAC ATCCTTATGT
551 GTAGTATATC CTCTAGATTT TGCCCGAACC CGATTAGGTG TCGATATTGG
601 AAAAGGTCTT GAGGAGCGAC AATTCAAGGG TTTAGGTGAC TGTATTATGA
651 AAATAGCAAA ATCAGATGGA ATTGCTGGTT TATACCAAGG GTTTGGTGTT
701 TCAGTACAGG GCATCATTGT GTACCGAGCC TCTTATTTTG GAGCTTATGA
751 CACAGTTAAG GGTTTATTAC CAAAGCCAAA GAAAACCTCA TTTCTTGTCT
801 CCTTTTTCAT TGCTCAAGTT GTGACTACAT GCTCTGGAAT ACTTCTTAT
851 CCGTTTGACA CAGTTAGAAG ACGTATGATG ATGCAGAGTG GTGAGGCTAA
901 ACGGCAATAT AAAGGAACCT TAGACTGCTT TGTGAAGATA TACCAACATG
951 AAGGAATCAG TTCCTTTTTC CGTGGCGCCT TCTCCAATGT TCTTCGCGGT
1001 ACAGGGGGTG CTTTGGTGTT GGTATTATAT GATAAAATTA AAGAAATCTT
1051 TCATATTGAT ATTGGTGGTA GGTAAATCGG AGAGTAAATT AAGAAATAAC
1101 ATGGATTTAA CTTGTTAAAC ATACAAATTA CATAGCTGCC ATTTGCATAC
1151 ATTTTGATAG TGTATTGTG TGTATTTTGT TAAAGTGCTA GTTCTGCAAT
1201 AAAGCATACA TTTTTCGAAG AATTAAATA CTAAATCA GATAAATGTG
1251 GATTTTCCCT CCACTTAGAC TCAACACAT TTTAGTGTGA TATTTTCATT
1301 ATTATAGGTA GTATATTTTA ATTTGTTAGT TTAATTTCT TTTTATGATT
1351 AAAAATTAAT CATATAATCC TAGATTAATG CTGAAATCTA GGAATGAAA
1401 GTAGCGTCTT TTAATTTGCT ATTCATTAA TATACCTGTT TTCCATCTT
1451 TTGAAGTCAT ATGGTATGAC ATATTTCTTA AAAGCTTATC AATAGATGTC
1501 ATCATATGTG TAGGCAGAAA TAAGCTTTGT TCTATATCTC TTCTAAGACA
1551 GTTGTATTAT CTGTGTATAA TATTACAGT ATCAGCCTTT GATTATAGAT
1601 GTGATCATTT AAAATTTGAT AATGACTTTA GTGACATTAT AAAACTGAAA
1651 CTGGAAAATA AAATGGCTTA TCTGCTGATG TTTATCTTTA AAATAAATAA
1701 AATCTTGCTA GTGTGAATAT ATCTTAGAAC AAAAGGTATC CTCTTGAAAA
1751 TTAGTTTGTA TATTTTGTG ACAATAAAGG AAGCTTAACT GTTAAAAAAA
1801 AAA
```

BLAST Results

No BLAST result

Medline entries

96289608:

Molecular biological and quantitative abnormalities of
ADP/ATP carrier protein in cardiomyopathic hamsters.

Peptide information for frame 2

ORF from 128 bp to 1072 bp; peptide length: 315
 Category: strong similarity to known protein
 Classification: Metabolism
 Prosite motifs: MITOCH CARRIER (40-50)
 MITOCH_CARRIER (145-155)
 MITOCH_CARRIER (242-252)

```

1 MHREPAKKKA EKRLFDASSF GKDLLAGGVA AAVSKTAVAP IERVKLLQV
51 QASSKQISPE ARYKGMVDCL VRIPREQGFF SFWRGNLANV IRYFPTQALN
101 FAFKDKYKQL FMSGVNEKEQ FWRWFLANLA SGGAAGATSL CVVYPLDFAR
151 TRLGVDIGKG PEERQFKGLG DCIMKIAKSD GIAGLYQGFG VSVQGIIVYR
201 ASYFGAYDTV KGLLPKPKKT PFLVSFFIAQ VVTTCSGILS YPFDTVRRRM
251 MMQSGEAKRQ YKGTLD CFVK IYQHEGISSF FRGAFSNVLR GTGGALVLVL
301 YDKIKEFFHI DIGGR

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_35n12, frame 2

PIR:S37210 ADP,ATP carrier protein T1 - mouse, N = 1, Score = 1127, P = 2.7e-114

PIR:A44778 ADP,ATP carrier protein T1 - human, N = 1, Score = 1125, P = 4.4e-114

TREMBL:DMADPATPT 2 product: "ADP/ATP translocase"; Drosophila melanogaster gene encoding ADP/ATP translocase, N = 1, Score = 1124, P = 5.6e-114

PIR:XWBO ADP,ATP carrier protein T1 - bovine, N = 1, Score = 1121, P = 1.2e-113

>PIR:S37210 ADP,ATP carrier protein T1 - mouse
 Length = 298

HSPs:

Score = 1127 (169.1 bits), Expect = 2.7e-114, P = 2.7e-114
 Identities = 214/293 (73%), Positives = 248/293 (84%)

```

Query:   17 ASSFGKDLLAGGVAAAVSKTAVAPIERVKLLQVQASSKQISPEARYKGMVDCLVRIPRE 76
          A SF KD LAGG+AAAVSKTAVAPIERVKLLQVQ +SKQIS E +YKG++DC+VRIP+E
Sbjct:   5 ALSFLKDFLAGGIAAAVSKTAVAPIERVKLLQVQHASKQISAQKQYKGIIDCVVRIPKE 64

Query:   77 QGFFSFWRGNLANVIRYFPTQALNFAFKDKYKQLFMSGVNEKEQFWRWFLANLASGGAAG 136
          QGF SFWRGNLANVIRYFPTQALNFAFKDKYKQ+F+ GV++ KQFWR+F NLASGGAAG
Sbjct:   65 QGFLSFWRGNLANVIRYFPTQALNFAFKDKYKQIFLGGVDRHKQFWRVFAGNLASGGAAG 124

Query:   137 ATSLCVVYPLDFARTRLGVDIGKGPEERQFKGLGDCIMKIAKSDGIAGLYQGFGVSVQGI 196
          ATSLC VYPLDFARTRL D+GKG +R+F GLGDC+ KI KSDG+ GLYQGF VSVQGI
Sbjct:   125 ATSLCFVYPLDFARTRLAADVKGSSQREFNGLGDCCLKIFKSDGLKGLYQGFSVSVQGI 184

Query:   197 IVYRASYFGAYDTVKGLLPKPKKTPFLVSFFIAQVVTTCSGILSYPFDTVRRRMMMQSGE 256
          I+YRA+YFG YDT KG+LP PK +VS+ IAQ VT +G++SYPFDTVRRRMMMQSG
Sbjct:   185 IIYRAAYFGVYDTAKGMLPDPKNVHIIVSWMIAQSVTAVAGLVSYPFDTVRRRMMMQSGR 244

Query:   257 --AKRQYKGTLD CFVKIYQHEGISSFFRGAFSNVLRGTGGALVLVLYDKIKEF 307
          A Y GTLDC+ KI + EG ++FF+GA+SNVLRG GGA VLVLYD+IK++
Sbjct:   245 KGADIMYTGTLDCWRKIAKDEGANAFKGAWSNVLRGMGGAFLVLVLYDEIKKY 297

```

Pedant information for DKFZphtes3_35n12, frame 2

Report for DKFZphtes3_35n12.2

[LENGTH] 315

```
SEQ      MHREPAKKKAEKRLFDASSFGKDLLAGGVAAAVSKTAVAPIERVKLLQVQASSQISPE
SEG
PRD      cccchhhhhhhhhhhhchhhhhhhhhchhhhhhhhhcchhhhhhhhhhhhhhhhhhhhh
MEM      .....

SEQ      ARYKGMVDCLVRIPREQGFFSFWRGNLANVIRYFPTQALNFAFKDKYKQLFMGSVNKEQ
SEG
PRD      hhhhhhhhheeeccccceeeeccccceeeecccchhhhhhhhhhhhhhhcccccccc
MEM      .....

SEQ      FWRWFLANLASGGAAGATSLCVVYPLDFARTRLGVDIGKGPEERQFKGLGDCIMKIAKSD
SEG      .... xxxxxxxxxxxxxxxxx
PRD      eeeeeccccccccccceeeeecchhhhhhhhhhhccccchhhhhhhcccceeeeeccc
MEM      .....

SEQ      GIAGLYQGFGVSQGIIVRYASYFYGAYDTVKGLLPKPCKTPFLVSFFIAQVVTTCSGILS
SEG
PRD      cccccccccceeeehhhhhccccccccccccccccccccchhhhhhhhhhhheeeec
MEM      ...MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM...

SEQ      YPFDTVRRRMMMQSGEAKRQYKGTLDCEVKIYQHEGISSFFRGAFSNVLRGTGGALVLVL
SEG
PRD      cccchhhhhhhhhccccceeeecccchhhhhhhhhccccccccccchhhhhccccceeeee
MEM      MCMMMMMMMMMM.....

SEQ      YDKIKEFFHIDIGGR
SEG
PRD      hhhhhhhhheeecccc
MEM
```

Prosite for DKFZphtes3_35n12.2

PS00215	40->50	MITOCH_CARRIER	PDOC00189
PS00215	145->155	MITOCH_CARRIER	PDOC00189
PS00215	242->252	MITOCH_CARRIER	PDOC00189

Pfam for DKFZphtes3_35n12.2

HMM_NAME	Mitochondrial carrier proteins		
HMM	*pFwkdfLAGGIAGmMeHTvMFPIDtIKTRMQlQgEMpM..ahpRYkGMI		
	+F+KD+LAGG+A++++T+++PI+++K+++Q+Q +++ RYKGM+		
Query	19	SFGKDLLAGGVAAVSKTAVAPIERVKLLQVQASSKQISPEARYKGMV	67
HMM	dCFRwIwkNEGWRGLWRGLgANvIRYIPqWaIRFGFYEFMKemFiDyfce		
	DC+ +I++++G++++WRG++ANVIRY+P++A++F+F++ +K +F + +++		
Query	68	DCLVRIPREQGFFSFWRGNLANVIRYFPTQALNFAFKDKYKQLEMSGVNK	117
HMM	ddnyWmWFwmnYMaGsmAGEwisvIitYPMWvVKTRLQaDqkHphsQp.R		
	++W+WF+ N+++G++AG ++S+ ++YP+++++TRL D +++++ R		
Query	118	EKQFWRWFLANLASGGAAG-ATSLCVVYPLDFARTRLGVD--IGKGPEER	164
HMM	hYNGvWNCwrkIYReEGgFkGLYRGWtPTWMMRMIpYqmiYffvYEtLkew		
	+++G+ +C KI +++G ++GLY+G++ +++++I+Y++ YF++Y+T K +		
Query	165	QFKGLGDCIMKIAKSDG-IAGLYQGFGVSVQGIIVYRASYFGAYDTVKGL	213
HMM	lynYtgYnPgprelCMddsPwWhWiIgWmIAGMiaWivSYpFDVVRTRMM		
	L +++ + ++++++I++ ++ +++++I+SYpFD+VR+RMM		
Query	214	LP-----KPK--KTPFLVSFFIAQVVT-TCSGILSYpFDTVRRRMM	251
HMM	Mdsm.edhkYqSmlDCWMqIYKnEGFkGEWKGFWPRIMRiMPWtAIMFmI		
	M+S+ +++++Y+++LDC+++IY++EG+ +F++G+ +++R+ ++A+++++		
Query	252	MQSGEAKRQYKGTLDcfVKIYQHEGISSFFRGAFSNVLRGT-GGALVLVL	300
HMM	YEqMKwFL*		
	Y+ +K+F+		
Query	301	YDKIKEFF	308

DKFZphtes3_35n24

group: testes derived

DKFZphtes3_35n24 encodes a novel 365 amino acid protein without similarity to known proteins.

The novel protein contains a Prosite Ig(Immunoglobulin)-MHC pattern. This pattern represents a domain, approximately one hundred amino acids long and including a conserved intra-domain disulfide bond (Ig domain). Thus, the novel protein is a new member of the Ig-superfamily. No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 1589 bp

Poly A stretch at pos. 1579, polyadenylation signal at pos. 1560

```
1 CGATCGTCAC GTGACGCCGG GGTTCAGCGT ATCCTTGCTG GGCAACCGTC
51 TTAGAGACCA GCACTGCTGG CTGCACCATG AATGTGATCT ACCCACTGGC
101 AGTCCCCAAG GGGCGCAGAC TCTGCTGTGA GGTGTGCGAA GCCCCAGCCG
151 AGCGGGGTGTG CGCGGCCTGC ACAGTCACTT ATTACTGTGG GGTGGTACAT
201 CAGAAAGGCTG ACTGGGACAG CATCCATGAG AAAATATGTC AGCTCTTGAT
251 TCCACTGCGC ACTTCCATGC CCTTCTACAA TTCAGAGGAA GAACGGCAGC
301 ATGGCCTGCA GCAGCTGCAG CAGCGGCAGA AGTATTGAT TGAATCTGTC
351 TACACCATAG CCCAGAAATA CCTCTTTGAA GGGAAACACG AAGATGCTGT
401 ACCAGCAGCT TTGCAGTCCC TTCGCTTCCG TGTGAAGCTG TATGGCCTGA
451 GCTCCGTAGA GCTTGTGCCT GCTTACCCGC TGTGGCCGA GGCCAGCCTT
501 GGTCTGGGCC GAATCGTTCA GGCTGAAGAA TATCTATTCC AAGCCCAAGT
551 GACAGTCCTC AAATCAACTG ACTGTAGTAA TGCCACCCAC TCTTTACTGC
601 ATCGGAATCT GGGACTTCTC TATATAGCTA AGAAAACTA TGAAGAGGCC
651 CGTTATCATC TGGCCAATGA TATTTATTTT GCCAGTTGTG CATTTGGAAC
701 AGAGGACATT AGGACTTCAG GAGGCTACTT CCACCTGGCT AATATATTCT
751 ATGACCTTAA AAAGTTGGAC CTGGCAGACA CATTGTACAC CAAGGTCTCT
801 GAGATCTGGC ATGCATATTT GAACAATCAC TATCAAGTCC TCTCACAGGC
851 TCACATCCAA CAAATGGATT TACTGGGCAA ACTATTGAG AATGACACTG
901 GCTTGGATGA AGCCCAAGAA GCAGAAGCCA TTCGCATCCT GACTTCAATC
951 TTGAACATTC GAGAATCTAC ATCTGACAAA GCCCCCCAAA AAACCATCTT
1001 TGTTCGTAAG ATCCTGGTCA TGCTTTACTA CCTGATGATG AATTCTTCAA
1051 AGGCACAGGA ATATGGCATG AGGGCCCTCA GTCTAGCCAA AGAACAACAG
1101 CTTGATGTCC ATGAGCAAAG CACCATTCAA GAGTTATTAA GTCTCATTTT
1151 AACTGAAGAC CATCCCATTA CTTAGTGACC CATGAGCTCT GCATCAAGGG
1201 TTATTCCAGG GGCTACTGAA GATCTAATAT ATTCCAGCCT TGCACAACCTG
1251 CTTTGAGGTA CTGTAGACTG CTGAAGTTTC CACCTCTTTC CCCTGGGATT
1301 GCACACATAG CTGTTATTTT TTTCTTACAC AGCATATTAA GGGAAATATA
1351 AGCTTTAGGC ATAGAAATCA CTAAAACTG TGTTTGTCAT GACCTTTGTA
1401 CTTGATTTAT CATGACTTTG TATGACTGAG TAATATGTAG TCAGATCACT
1451 AATATGGTAT TTGTAATTAA ACTACAAATA GTTTGTCATT TCCCAGAAGT
1501 CTTCAACGA TGCAATGTTT ATACACTTTT GCTAAAGGAG GGGTAAAGGA
1551 GGGGGTAGGG AATAAAGCTA TATTGGAACA AAAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 78 bp to 1172 bp; peptide length: 365
Category: putative protein

Prosite motifs: IG_MHC (35-42)

```

1 MNVIYPLAVP KGRRLCCEVC EAPAERVCAA CTVTYCGVV HQKADWDSIH
51 EKICQLLIPL RTSMPFYNSE EERQHGLQOL QQRQKYLIEF CYTIAQKYLE
101 EKGHEDAVPA ALQSLRFRVK LYGLSSVELV PAYPLLAES LGLGRIVQAE
151 EYLFQAQWTV LKSTDCSNAT HSLLRNLGL LYIAKKNYEE ARYHLANDIY
201 FASCAFGTED IRTSGGYFHL ANIFYDLKKL DLADTLTKV SEIWHAYLNN
251 HYQVLSQAH IQQMDLLGKLF ENDTGLDEAQ EAEAIRILTS ILNIRESTSD
301 KAPQKTIFVL KILVMLYYLM MNSSKAQEGY MRALSLAKEQ QLDVHEQSTI
351 QELLSLISTE DHPIT

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_35n24, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_35n24, frame 3

Report for DKFZphtes3_35n24.3

```

[LENGTH]      365
[MW]           41768.24
[pI]           5.82
[BLOCKS]      BL00273 Heat-stable enterotoxins proteins
[PROSITE]     MYRISTYL 1
[PROSITE]     IG_MHC 1
[PROSITE]     AMIDATION 1
[PROSITE]     CK2_PHOSPHO_SITE 7
[PROSITE]     TYR_PHOSPHO_SITE 4
[PROSITE]     PKC_PHOSPHO_SITE 3
[PROSITE]     ASN_GLYCOSYLATION 3
[KW]          Alpha Beta
[KW]          LOW_COMPLEXITY 4.11 %

```

```

SEQ  MNVIYPLAVPKGRRLCCEVCEAPAERVCAACTVTYTCGVVHQKADWDSIHEKICQLLIPL
SEG  .....
PRD  cccceeeccccceeeeeeheehhhhhhhheeeeeeeccccccccchhhhhhhheec

SEQ  RTSMPFYNSEERQHGLQOLQQRQKYLIEFCYTIAQKYLEFEGKHEDAVPAALQSLRFRVK
SEG  .....xxxxxxxxxxxxxxxx
PRD  cccccccchhhhhhhchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccchhhhhhhhhhh

SEQ  LYGLSSVELVPAYPLLAESLGLGRIVQAEYLFQAQWTVLKSTDCSNATHSLLHRNLGL
SEG  .....
PRD  hhccccceccccchhhhhccccchhhhhhhhhhhhhhhhhhhhhcccccccccccccccccccc

SEQ  LYIAKKNYEARYHLANDIYFASCAFGTEDIRTSGGYFHLANIFYDLKKLADTLTKV
SEG  .....
PRD  eeeehhhhhhhhhhhhhheeeccccccccccccceehhhhhhhhhhhhhccccceeeeh

SEQ  SEIWHAYLNNHYQVLSQAH IQQMDLLGKLFENDTGLDEAQEAEAIRILTSILNIRESTSD
SEG  .....
PRD  hhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccc

SEQ  KAPQKTIFVLKILVMLYYLMNSSKAQEGYMRALSLAKEQQLDVHEQSTIQELLSLISTE
SEG  .....
PRD  cccccceeehhhhhhhhhhhhhhccccchhhhhhhchhhhhhhhhhhhhhhhhhhhhhhhhhhhhcc

SEQ  DHPIT
SEG  .....
PRD  ccccc

```

Prosite for DKFZphtes3_35n24.3

PS00001	168->172	ASN_GLYCOSYLATION	PDOC00001
PS00001	272->276	ASN_GLYCOSYLATION	PDOC00001
PS00001	322->326	ASN_GLYCOSYLATION	PDOC00001
PS00005	114->117	PKC_PHOSPHO_SITE	PDOC00005
PS00005	299->302	PKC_PHOSPHO_SITE	PDOC00005
PS00005	323->326	PKC_PHOSPHO_SITE	PDOC00005

PS00006	48->52	CK2_PHOSPHO_SITE	PDOC00006
PS00006	69->73	CK2_PHOSPHO_SITE	PDOC00006
PS00006	125->129	CK2_PHOSPHO_SITE	PDOC00006
PS00006	274->278	CK2_PHOSPHO_SITE	PDOC00006
PS00006	297->301	CK2_PHOSPHO_SITE	PDOC00006
PS00006	349->353	CK2_PHOSPHO_SITE	PDOC00006
PS00006	358->362	CK2_PHOSPHO_SITE	PDOC00006
PS00007	85->93	TYR_PHOSPHO_SITE	PDOC00007
PS00007	186->194	TYR_PHOSPHO_SITE	PDOC00007
PS00007	186->194	TYR_PHOSPHO_SITE	PDOC00007
PS00007	185->194	TYR_PHOSPHO_SITE	PDOC00007
PS00008	275->281	MYRISTYL	PDOC00008
PS00009	11->15	AMIDATION	PDOC00009
PS00290	35->42	IG_MHC	PDOC00262

(No Pfam data available for DKFZphtes3_35n24.3)

DKFZphtes3_35n9

group: metabolism

DKFZphtes3_35n9 encodes a novel 607 amino acid protein which is a splice variant of human carboxylesterase (EC 3.1.1.1).

The novel protein contains both, one carboxylesterase B1 and one B2 pattern. In comparison to EC 3.1.1.1, DKFZphtes3_35n9 shows a N-terminal extension and aa 458-474 are missing.

The new protein can find application in modulation of carboxylester metabolism and as a new enzyme for biotechnologic production processes.

carboxylesterase, splice variant

5' extension of mRNA and N-terminal elongation of protein (64 aa), missing exon! aa 458-474 of JC5408 are missing

Sequenced by DKFZ

Locus: unknown

Insert length: 2888 bp

Poly A stretch at pos. 2878, no polyadenylation signal found

```

1 CTCGGCCTGA GGTGCGAGAG AAGCGGTGAC CGCGGCCCTG GCTGCTCGGA
51 CCCGGGAACA TGATGGTCGC TGGAGCAGAA GGCGCTGAGA AGGGACCACG
101 GCGGCGCTGG GTCGTGCAGG CCAGTAGCGG GCTGAAACGT AGAGGCCAGA
151 ACCAGGTCTC AGGGGGCACT AAAGGCGGTC GGAGGTAATC CCCACACCGC
201 TTCTCTCTGG AAGTCAGGCT GGCCGGGAGC TCCCGTATCC AGGACGGTTG
251 GTCGCCTCTG GCCTGGCAGG GATCCTAGTG TCTCGGGACC TCCCGGTGAC
301 GCGCCTGCCT CCCCTGCTGC ACCATAGGCC CGGGAGTACG GCGTCCCCAC
351 AGCTTGGACC GGCAGGGGCT CGTGAAATGT TTGTCAAGTG GATAAATGAC
401 CATGGCCGTG GTCTCCGCGG GAGGTGAGGA AACTGAAAGC CACCGAGGAA
451 AAGGGGGGCG CTCCTTAAGA AGTGCCGCGG TCACGTGTAC GTTTCAAAAG
501 AATGGCGTGA CTGAGTAGGG AGGGGACCGC GGAGACCTTC AGACCCTGGA
551 CTGTAAGGAG ATGAGGGGCC GTGAAGGGGA ACCCAGGAAA CTGAGTCCTG
601 AAAGCAAGGA GGAACCTCCA GAATGAAGGG CGCCGACACT CCTTCTTGCC
651 TTTGCTCAAG CGTTCTCTTC ACCCCGATCA AGTTCCTTCC CATTTCTCCA
701 TCTGGGGGAT CCTGAACGTG CACATCCTCA GAGAAGCCCT CCTGGGGTCT
751 CCAATTCTAG TTTATTGCCC CCTCCTATCG ATCCCCCAGC GCGCTCATCG
801 GGCTGTGGA CAAGGACAGG TTTGAAGAGA GGATTCCTTG GATCGCGGAA
851 GGGCTGCAGG AATGGCACAG CCCCTTCCGA GGATGCCAAA GGAGCCCGGG
901 CAAAGGAAAG TGGCCGTGCC CGGGCCTGCC TACCACTAGA TCCCCACCA
951 CCTATGACTG CTCAGTCCCG CTCCTCCTACC ACACCCACCT TTCCCGGCCC
1001 AAGCCAGCGC ACCCCGCTGA CTCCTTGCCC AGTCCAACT CCAAGGCTGG
1051 GCAAGGCACT GATCCACTGC TGGACAGACC CGGGGACAGC TCTGGGTGAA
1101 CAGCAGCGTG TCCGCCGCGA GCGAACCGAG ACCAGCGAGC CGACCATGCG
1151 GCTGCACAGA CTTGCTGCGC GGCTGAGCGC GGTGGCCTGT GGGCTTCTGC
1201 TGCTTCTTGT CCGGGGCCAG GGCAGGACT CAGCCAGTCC CATCCGGACC
1251 ACACACACGG GGCAGGTGCT GGGGAGTCTT GTCCATGTGA AGGGCGCCAA
1301 TGCCGGGGTG CAAACCTTCC TGGGAATTCC ATTTGCCAAG CCACCTCTAG
1351 GTCCGCTGCG ATTTGCACCC CCTGAGCCCC CTGAATCTTG GAGTGGTGTG
1401 AGGGATGGAA CCACCCATCC GGCGATGTGT CTACAGGACC TCACCGCAGT
1451 GGAGTCAGAG TTTCTTAGCC AGTTCAACAT GACCTTCCCT TCCGACTCCA
1501 TGTCTGAGGA CTGCCTGTAC CTCAGCATCT ACACGCCGGC CCATAGCCAT
1551 GAAGGCTCTA ACCTGCCGGT GATGGTGTGG ATCCACGGTG GTGCGCTTGT
1601 TTTTGGCATG GCTTCTTGT ATGATGGTTC CATGCTGGCT GCCTTGAGGA
1651 ACGTGGTGGT GGTGTCATC CAGTACCGCC TGGGTGTCTT GGGCTTCTTC
1701 AGCACTGGAG ACAAGCACGC AACCGGCAAC TGGGGCTACC TGGACCAAGT
1751 GGCTGCACTA CGCTGGGTCC AGCAGAATAT CGCCCACTTT GGAGGCAACC
1801 CTGACCGTGT CACCATTITT GCGGAGTCTG CGGGTGGCAC GAGTGTGTCT
1851 TCGCTTGTG TGTCCCCAT ATCCCAAGGA CTCTTCCACG GAGCCATCAT
1901 GGAGAGTGGC GTGGCCCTCC TGCCCGGCTT CATTGCCAGC TCAGCTGATG
1951 TCATCTCCAC GGTGGTGGCC AACCTGTCTG CCTGTGACCA AGTTGACTCT
2001 GAGGCCCTGG TGGGCTGCCT GCGGGGCAAG AGTAAAGAGG AGATTCTTGC
2051 AATTAACAAG CCTTCAAGA TGATCCCGG AGTGGTGGAT GGGGTCTTCC
2101 TGCCAGGCA CCCCAGGAG CTGCTGGCCT CTGCGGACTT TCAGCCTGTC
2151 CCTAGCATTG TTGGTGTCAA CAACAATGAA TTCGGCTGCG TCATCCCCAA
2201 GGTGATGAGG ATCTATGATA CCCAGAAGGA AATGGACAGA GAGGCCTCCC
2251 AGGCTGCTCT GCAGAAAATG TTAACGCTGC TGATGTTGCC TCCTACATTT
2301 GGTGACCTGC TGAGGGAGGA GTACATTGGG GACAATGGGG ATCCCCAGAC
2351 CCTCCAAGCG CAGTTCAGG AGATGATGGC GGAATCCATG TTTGTGATCC
2401 CTGCACTCCA AGTAGCACAT TTTCACTGTT CCCGGGCCCC TGTGACTTTC
2451 TACGAGTTC AGCATCAGCC CAGCTGGCTC AAGAACATCA GGCCACCGCA
2501 CATGAAGGCA GACCATGTTA AATTCACTGA GGAAGAGGAG CAGCTAAGCA
2551 GGAAGATGAT GAAGTACTGG GCCAACTTTG CGAGAAATGG GAACCCCAAT
2601 GGGGAGGGTC TGCCACACTG GCCGCTGTTC GACCAGGAGG AGCAATACCT

```

```

2651 GCAGCTGAAC CTACAGCCTG CGGTGGGCGG GGCTCTGAAG GCCCACAGGC
2701 TCCAGTTCTG GAAGAAGGCG CTGCCCCAAA AGATCCAGGA GCTCGAGGAG
2751 CCTGAAGAGA GACACACAGA GCTGTAGCTC CCTGTGCCGG GGAGGAGGGG
2801 GTGGGTTCGC TGACAGGCGA GGGTCAGCCT GCTGTGCCCA CACACACCCA
2851 CTAAGGAGAA AGAAGTTGAT TCCTTCATAA AAAAAAAA

```

BLAST Results

Entry D50579 from database EMBL:
Homo sapiens mRNA for carboxylesterase, complete cds.
Score = 7197, P = 0.0e+00, identities = 1441/1443

Entry JC5408 from database PIR:
carboxylesterase (EC 3.1.1.1) - human
Score = 2808, P = 1.2e-291, identities = 542/559, positives = 543/559,
frame +3

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 954 bp to 2774 bp; peptide length: 607
Category: known protein
Classification: Metabolism
Prosite motifs: CARBOXYLESTERASE_B_1 (279-295)
CARBOXYLESTERASE_B_2 (185-196)

```

1 MTAQSRSPPT PTFPGPSQRT PLTPCPVQTP RLKALIHCV TDPGQPLGEQ
51 QRVRQRQRTT SEPTMLRLHRL RARLSAVACG LLLLLVVRGQ QDSASPIRTT
101 HTGQVLGSLV HVKGANAGVQ TFLGIPFAKP PLGPLRFAPP EPPESWSGVR
151 DGTTHPAMCL QDLTAVESEF LSQFNMTFPS DSMSEDCLYL SIYTPAHSHE
201 GSNLPVMVWI HGGALVFGMA SLYDGSMLAA LENVVVVIQ YRLGVLGFFS
251 TGDKHATGNW GYLDQVAALR WVQQNIAHEG GNPDRVITFG ESAGGTSVSS
301 LVVSPISQGL FHGAIMESGV ALLPGLIASS ADVISTVVAN LSACDQVDSE
351 ALVGCLRGKS KEEILAINKP FKMIPGVVDG VFLPRHPQEL LASADFQVPV
401 SIVGVNNNEF GWLIPKVMRI YDTQKEMDRE ASQAALQKML TLLMLPPTFG
451 DLLREEYIGD NGDPQTLQAO FQEMMADSMF VIPALQVAHF QCSRAPVYFY
501 EFQHQPSWLK NIRPPHMKAD HVKFTEEEEQ LSRKMMKYWA NFARNGNPNG
551 EGLPHWPLFD QEEQYLQLNL QPAVGRALKA HRLQFWKKAL PQKIQEELEP
601 EERHTEL

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_35n9, frame 3

PIR:JC5408 carboxylesterase (EC 3.1.1.1) - human, N = 1, Score = 2808,
P = 1.9e-292

TREMBL:HSU60553_1 gene: "hCE-2"; product: "carboxylesterase"; Human
carboxylesterase (hCE-2) mRNA, complete cds., N = 1, Score = 2761, P =
1.8e-287

PIR:A34329 60K esterase (EC 3.1.1.-) isoform 2 - rabbit, N = 1, Score =
1985, P = 3.1e-205

TREMBL:D50580_1 product: "carboxylesterase precursor"; Rattus
norvegicus mRNA for carboxylesterase, partial cds., N = 1, Score =
1984, P = 4e-205

>PIR:JC5408 carboxylesterase (EC 3.1.1.1) - human
Length = 559

HSPs:

Score = 2808 (421.3 bits), Expect = 1.9e-292, P = 1.9e-292

Identities = 542/559 (96%), Positives = 543/559 (97%)

Query: 65 MRLHRLRRLSAVACGLLLLLVRGQGDSPARTTHTGQVLGSLVHVKGANAGVQTFLG 124
 Sbjct: 1 MRLHRLRRLSAVACGLLLLLVRGQGDSPARTTHTGQVLGSLVHVKGANAGVQTFLG 60

Query: 125 IPFAKPLGLPLRFAPPEPPESWSGVRDGTTHPAMCLQDLTAVESEFLSQFNMTFPSDSMS 184
 Sbjct: 61 IPFAKPLGLPLRFAPPEPPESWSGVRDGTTHPAMCLQDLTAVESEFLSQFNMTFPSDSMS 120

Query: 185 EDCLYLSIYTPAHSHEGSNLPVMVWIHGGALVFGMASLYDGSMALAALENVVVVIIQYRLG 244
 Sbjct: 121 EDCLYLSIYTPAHSHEGSNLPVMVWIHGGALVFGMASLYDGSMALAALENVVVVIIQYRLG 180

Query: 245 VLGFFSTGDKHATGNWGYLDQVAALRWVQQNIAHFGGNPDRVTIFGESAGGTSVSSSLVVS 304
 Sbjct: 181 VLGFFSTGDKHATGNWGYLDQVAALRWVQQNIAHFGGNPDRVTIFGESAGGTSVSSSLVVS 240

Query: 305 PISQGLFHGAIMESGVALLPGLIASSADVISTVVANLSACDQVDSEALVGCRLGKSKEEI 364
 Sbjct: 241 PISQGLFHGAIMESGVALLPGLIASSADVISTVVANLSACDQVDSEALVGCRLGKSKEEI 300

Query: 365 LAINKPFKMI PGVVDGVFLPRHPQELLASADFQPVPSIVGVNNNEFGWLIPKVMRIYDTQ 424
 Sbjct: 301 LAINKPFKMI PGVVDGVFLPRHPQELLASADFQPVPSIVGVNNNEFGWLIPKVMRIYDTQ 360

Query: 425 KEMDREASQAALQKMLTLLMLPPTFGDLLREEYIGDNGDPQTLQAQFQEMMADSMFVIPA 484
 Sbjct: 361 KEMDREASQAALQKMLTLLMLPPTFGDLLREEYIGDNGDPQTLQAQFQEMMADSMFVIPA 420

Query: 485 LQVAHFQCSRAPVYFYEFOHQPSWLKNIRPPHMKADH-----VKFTEEE 528
 Sbjct: 421 LQVAHFQCSRAPVYFYEFOHQPSWLKNIRPPHMKADH +KFTEEE 480

Query: 529 EQLSRKMMKYWANFARNGNPNGEGLPHWPLFDQEEQYLQNLQPAVGRALKAHRLQFWKK 588
 Sbjct: 481 EQLSRKMMKYWANFARNGNPNGEGLPHWPLFDQEEQYLQNLQPAVGRALKAHRLQFWKK 540

Query: 589 ALPQKIQELEEPEERHTEL 607
 Sbjct: 541 ALPQKIQELEEPEERHTEL 559

Pedant information for DKFZphtes3_35n9, frame 3

Report for DKFZphtes3_35n9.3

[LENGTH] 607
 [MW] 67051.20
 [pI] 6.11
 [HOMOL] PIR:JC5408 carboxylesterase (EC 3.1.1.1) - human 0.0
 [BLOCKS] BL01173A Lipolytic enzymes "G-D-X-G" family, histidine
 [BLOCKS] BL00122G
 [BLOCKS] BL00122F
 [BLOCKS] BL00122E
 [BLOCKS] BL00122D Carboxylesterases type-B serine proteins
 [BLOCKS] BL00122C Carboxylesterases type-B serine proteins
 [BLOCKS] BL00122B Carboxylesterases type-B serine proteins
 [BLOCKS] BL00122A Carboxylesterases type-B serine proteins
 [SCOP] dlakn_ 3.56.1.1.4 Bile-salt activated lipase [Bovine (Bos taurus 1e-158
 [SCOP] d2ack_ 3.56.1.1.1 Acetylcholinesterase [Electric ray (Torped 1e-170
 [SCOP] dlthg_ 3.56.1.9.7 type-B carboxylesterase/lipase [fungu 1e-149
 [EC] 3.1.1.13 Sterol esterase 1e-52
 [EC] 3.1.1.17 Acetylcholinesterase 5e-74
 [EC] 3.1.1.1 Carboxylesterase 0.0
 [EC] 3.1.1.8 Cholinesterase 5e-68
 [EC] 3.1.1.59 Juvenile-hormone esterase 1e-34
 [EC] 3.1.1.3 Triacylglycerol lipase 3e-52
 [PIRKW] duplication 2e-47
 [PIRKW] homotetramer 3e-67
 [PIRKW] transmembrane protein 9e-44
 [PIRKW] microsome 1e-130
 [PIRKW] pancreas 3e-52
 [PIRKW] endoplasmic reticulum 1e-134
 [PIRKW] homotrimer 1e-134
 [PIRKW] phosphatidylinositol linkage 5e-74
 [PIRKW] synapse 3e-73
 [PIRKW] liver 1e-131
 [PIRKW] heparin binding 3e-52

```

[PIRKW]      phosphoprotein 7e-25
[PIRKW]      glycoprotein 1e-134
[PIRKW]      thyroid hormone biosynthesis 2e-47
[PIRKW]      carboxylic ester hydrolase 0.0
[PIRKW]      monomer 2e-42
[PIRKW]      disulfide bond 2e-31
[PIRKW]      mammary gland 3e-52
[PIRKW]      alternative splicing 5e-74
[PIRKW]      iodine 2e-47
[PIRKW]      pyroglutamic acid 6e-39
[PIRKW]      hydrolase 1e-135
[PIRKW]      muscle 3e-73
[PIRKW]      thyroid gland 2e-47
[PIRKW]      membrane protein 3e-73
[PIRKW]      neurotransmitter degradation 3e-73
[PIRKW]      cholesterol 3e-52
[PIRKW]      homodimer 2e-47
[PIRKW]      nerve 3e-73
[SUPFAM]     cholinesterase 0.0
[SUPFAM]     triacylglycerol lipase 1e-32
[SUPFAM]     cholinesterase homology 0.0
[SUPFAM]     thyroglobulin 2e-47
[SUPFAM]     thyroglobulin type I repeat homology 2e-47
[SUPFAM]     juvenile-hormone esterase 2e-35
[SUPFAM]     probable lipolytic protein ybaC 1e-07
[PROSITE]    CARBOXYLESTERASE_B_2 1
[PROSITE]    CARBOXYLESTERASE_B_1 1
[PFAM]       Carboxylesterases
[KW]         Alpha_Beta
[KW]         3D
[KW]         LOW_COMPLEXITY      3.95 %

SEQ      MTAQSRSPPTTPTFPGPSQRTPLTPCPVQTPRLGKALIHCVTDPGQPLGEQQRVRRQRTET
SEG      .....
lacj-    .....

SEQ      SEPTMRHLRLRLRLSAVACGLLLLLVRGQGQDSASPIRTTHTGQVLGSLVHVKGANAGVQ
SEG      .....
lacj-    .....ETEEEECEEEEEETTEE--EE

SEQ      TFLGIPFAKPLPLRFAPPEPPESWSGVDRGTTHPAMCLQDLTAVESEFLSQFNMTFPS
SEG      .....
lacj-    EEEEECEETTTGGGTTTCCEECCCCCEEECCCCCBCCCCCTTTTTT-HHHHHCCCC

SEQ      DSMSEDCLYLSIYTPAHSHEGSNLPVMVWIHGALVFGMASLYDGSMMLAALENVVVVVIQ
SEG      .....
lacj-    CCBTTTTCEEEEEET--TTTTTTEEEEEECTTTTTCTTTTGCHHHHHHHHCCEEEEECC

SEQ      YRLGVLGFFSTGDKHATGNWGYLDQVAALRWVQQNIAHFGGNPDRVTIFGESAGGTSVSS
SEG      .....
lacj-    CCCCCGGCCCTTTTTTTCCHHHHHHHHHHHHHHCGGGGCEEEEEEECHHHHHHHH

SEQ      LVVSPISQGLFHGAIMESGVALLPGLIASSADVISTVVANLSACDQVDSEALVGCLRGKS
SEG      .....
lacj-    HHHCGGGTTTTCCEEEETTTTTTTTTTBCHHHHHHHHHHHC-CCCCCHHHHHHHHHHC

SEQ      KEEILAINKPFKMPGVVDGVFLPRHPQELLASADFQPVPSIVGVNNNEFGWLI PKVMRI
SEG      .....
lacj-    HHHHHHHHTCCCTTTCBTTTTHHHHHHHHTTCCCCEEEEETBTHHHHHHTTTT

SEQ      YDTQKEMDREASQAALQKMLTLLMLPPTFGDLLREEYIGDNGDPQTLQAQFQEMMADSMF
SEG      .....
lacj-    TTTCCCCCHHHHHHHHHHHHTTTTCHHHHHHHHHHCTTTTTTHHHH-HHHHHHHHHHHH

SEQ      VIPALQVAHFQCSRAPVYFQHQPSWLKNIRPPHMKADHVKFTEEEEQLSRKMMKYWA
SEG      .....
lacj-    HHHHHHHHHHHHCCCEEEEECCCCGGGTBTTHHHCGGGCCCHHHHHHHHHHHHHH

SEQ      NFARNGPNNGEGLPHWPLFDQEEQYLQNLQPAVGRALKAHRLQFWKKALPKIQELEEPE
SEG      .....
lacj-    HHHHHCCCCCCC--CCCCBTTTBEEEECCCCCEETTHHHHHHHHHHHHH.....

SEQ      EERHTEL
SEG      xxxxxx.
lacj-    .....

```

Prosites for DKFZphtes3_35n9.3

PS00122 279->295 CARBOXYLESTERASE_B_1 PDOC00112
 PS00941 185->196 CARBOXYLESTERASE_B_2 PDOC00112

Pfam for DKFZphtes3_35n9.3

HMM_NAME	Carboxylesterases
HMM	*MfMnwlimFLlwmItWii.WhegaprpPdPyiVdtnnCGkIRGmNedtD + +L+++ ++++++ +Q+++P I T+ G + G ++ +
Query	69 RLRARLSAVACGLLLLLVRGQGQDSASP---IRTTHT-GQVLGSLVHVK 113
HMM	NG..pYYvFlGIPYAEPPVGNLRFKePQPYhePwtNVWNATnYPPMCMQW + + +FLGIP+A+PP+G LRF +P+P +E W++V++ T+ P MC+Q+
Query	114 GANAGVQTFLGIPFAKPLGLRFAPEP-PESWSGVRDGTTHPAMCLQD 162
HMM	ndFGFWlFdmieMWNeniP..eMSDECLYLNWVTPWnrkPNskLPVMVWI +++ ++N++ P +MSDECLYL+++TP+ + ++S+LPVMVWI
Query	163 LTAV--ESEFLSQFNMTFPSDSMSDECLYLSIYTPAHSHEGSNLPVMVWI 210
HMM	HGGGFMFGSGhsYPligYDgeylMeeNVIVVtINyRLGPFGLSTgDid HGG+++FG + ++YDG+ L++ ENV+VV I+YRLG++GF+STGD +
Query	211 HGGALVFGMA-----SLYDGSMALLENVVVVIQYRLGLVGFSTGDKH 255
HMM	lPPHGNWGLWDQRMALQWVQDNIANFGGDPNNITIFGESAGGMSVHlHML + GNWG++DQ++AL+WVQ+NIA+FGG+P+++TIFGESAGG+SV+ ++
Query	256 AT--GNWGYLDQVAALRWVQQNIAHFGGNPDRVTIFGESAGGTSVSSLVV 303
HMM	SYGGDNPPmfKqLFHRAIMQSGsAmcPWviQsnyNaRqRAFRFARimGCN S P + +LFH AIM+SG A+ P++I S++ + +A++ C+
Query	304 S-----PISQGLFHGAIMESGVALLPGLIASSA--DVISTVVANLSACD 345
HMM	rmDsseMIqCLRsKPwEELWdAtWnFwmWfyfPflPWFFgPVIDGDDaPE + DS++++ CLR K+ EE+++++ +F + + +DG+
Query	346 QVDSEALVGCLRGKSKEEILAINK----PFKMIPGV-----VDGV---- 381
HMM	aFIPDHPeMiKEGkFnDVPWIIGYNNDEGIWfapMmMfnWfdEDeWId F+P+HP+E++++ F VP I+G+NN E++W++P M + + +E++
Query	382 -FLPRHPQELLASADFQVPVSIVGVNNNEFGWLIPKVMRIYDT-QKEMDR 429
HMM	itNedWyeWMPYIlFYrddmsNikDMDDYiDkvyEeYpGWWDrfPqESYW ++ + ++ M +L + + + D ++EEY+G+ + PQ
Query	430 EASQAALQKMLTLLMLPPT-F-----GDLLREEYIGDNGD-PQTLQA 469
HMM	nLqDMFTDYLFWCpTRihadnHRkHwgsPVYMYeFDHPpSFGYgQFFmWR ++Q+M+ D F++P + ++H++ +PVY+YEF+H PS +
Query	470 QFQEMMADSMFVIP--ALQVAHFQCSRAPVYFYEFQHQPsw-----LKN 511
HMM	WWPpWMgvDH* +PP+M++DH
Query	512 IRPPHMKADH 521
HMM	*tEEeiissMRmMMNYWINFAKhGNPNnthnglCWWPqYTsneQYdMIMe TEEE+ +S R MM+YW+NFA++GNPN++ GL++WP +++EQY++ +
Query	525 TEEEEQLS-RKMMKYWANFARNGNPNGE--GLPHWPLFDQEEQYLQLNL 570
HMM	tIImiQmCrmrDPYCnFW* + +++++ + FW
Query	571 QPAVGRALKAHK--LQFW 586

DKFZphtes3_35p17

group: testes derived

DKFZphtes3_35p17 encodes a novel 505 amino acid protein with weak similarity to Proteins of the armadillo family.

Proteins of the armadillo family are involved in diverse cellular processes in higher eukaryotes. Some of them, like armadillo, beta-catenin and plakoglobins have dual functions in intercellular junctions and signalling cascades. Others, belonging to the importin-alpha-subfamily are involved in NLS recognition and nuclear transport, while some members of the armadillo family have as yet unknown functions. The novel protein shows similarity to S. cerevisiae protein Yel013p (VAC8) and Danio rerio b-catenin, but contains no armadillo (arm) repeats.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to S.cerevisiae VAC8

complete cDNA, complete cds, few EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 1966 bp

Poly A stretch at pos. 1956, polyadenylation signal at pos. 1935

```
1 AAGTCAAATG TAAGATTGGT TCATTAAAAA TACTGAAGGA AATCAGTCAT
51 AATCCTCAAA TCAGACAGAA TATTGTTGAC CTTGGGGGCT TACCAATTAT
101 GGTGAATATA CTTGATTCTC CACACAAGAG TCTAAAATGT TTGGCAGCCG
151 AGACTATCGC GAATGTTGCC AAGTTTAAAA GAGCACGGCG GGTGGTGAGG
201 CAGCACGGGG GTATCACCAA ACTGGTTGCT CTACTAGACT GTGCACATGA
251 TTCCACAAAA CTGCCCCAAT CGAGTCTGTA TGAGGCCAGA GACGTGGAAG
301 TGGCTCGCTG TGGGGCACTG GCCCTGTGGA GCTGCAGTAA GAGTCATACG
351 AATAAAGAAG CCATCCGCAA AGCTGGGGGC ATTCTCTGTG TGGCTCGGCT
401 GCTGAAGACT TCTCATGAAA ACATGCTAAT TCCAGTGGTG GGGACATTGC
451 AAGAGTGTGC ATCAGAGGAA AACTACCGGG CTGCAATCAA AGCAGAAAGG
501 ATCATTGAAA ACCTTGTCAA GAACCTAAAT AGTGAGAATG AGCAGCTGCA
551 GGAGCACTGC GCCATGGCCA TTTACCACTG TGCTGAAGAT AAGGAAACCC
601 GGGACCTCGT TAGGCTGCAC GGAGGACTTA AGCCCTTGCC CAGTCTACTC
651 AATAACACTG ACAATAAAGA GCGGTTAGCT GCTGTACAGG GGGCTATATG
701 GAAATGTTCC ATCAGCAAAG AGAATGTTAC CAAGTTTCGG GAATACAAAG
751 CCATTGAAAC CTTGGTGGGA CTTCTAACAG ATCAGCCTGA AGAAGTACTT
801 GTGAATGTGG TTGGGGCCTT GGGAGAATGC TGCCAAGAAG GTGAAAACCG
851 AGTCATTGTC CGGAAATGTG GTGGCATTCA ACCACTTGTG AACCTCCTTG
901 TTGGAATAAA CCAAGCTCTT CTTGTGAATG TTACAAAAGC AGTTGGTGCT
951 TGTGCAGTAG AACCTGAAAG TATGATGATA ATTGATCGCT TAGATGGAGT
1001 TCGTTTGTG TGGTCCCTGC TGAAAAATCC TCACCCAGAC GTGAAGGCCA
1051 GCGCAGCATG GGCACCTCTG CCATGCATCA AAAATGCAAA GGATGCTGGG
1101 GAAATGGTTC GTTCTTTTGT TGGTGGTTTG GAACTTATTG TCAATTTACT
1151 GAAATCAGAT AACAAAGAAG TTCTGGCAAG TGTATGTGCT GCCATTACCA
1201 ACATAGCAAA AGATCAAGAA AATTTAGCTG TTATCACAGA TCATGGAGTT
1251 GTTCTTTTAT TGTCCAAACT GGCAAAATCA AATAACAATA AATTGAGACA
1301 TCATCTAGCA GAAGCTATTT CACGTTGCTG TATGTGGGGC AGGAAATAGAG
1351 TGGCCTTCGG TGAGCACAAA GCAGTGGCTC CACTAGTGCG TTATCTGAAA
1401 TCAAATGACA CCAACGTGCA TCGGGCGACA GCTCAGGCCT TGTACCAACT
1451 CTCAGAAGAC GCCGATAACT GCATCACCAT GCATGAGAA TGGTGCAGTAA
1501 AGCTTCTACT GGATATGGTT GGGTCCCTG ACCAGGATCT CCAGGAAGCT
1551 GCAGCTGGTT GTATATCCAA TATCCGCAGG CTGGCTCTTG CTACAGAGAA
1601 GGCAAGATAC ACTTGAAATT TAAATGGACA TTACAAGCTA TCAAATTCTA
1651 CATGACACAG GACATGTCAC TCCCATGGCC AGAAAGCCTA AATTGGGAAA
1701 CAGTTGTTAG CAAACCCTTT CAACCATCTA AATGAAAACA CACAAATTGA
1751 AAATGCACAG AATGTTTTTC ATCTGAAAT TGCATGGAGA CTTTTGTTTC
1801 TATTTAATGT TTTCGAGATA TGACATGTGA TAAGATGGAA AGCCAATAAA
1851 CCTGTGATAA GTTTCTAAGA ATATGAGAA ATACGTATAT GATGTATTTT
1901 TAGTTCAGTG ATGCTTTTGT ATTTGTGGCG ATTTTAATAA AGGATATGGC
1951 CTTCCCAAAA AAAAAA
```

BLAST Results

No BLAST result

Medline entries

98413148:
Yel013p (Vac8p), an armadillo repeat protein related to plakoglobin and importin alpha is associated with the yeast vacuole membrane.

98330438:
YEB3/VAC8 encodes a myristylated armadillo protein of the *Saccharomyces cerevisiae* vacuolar membrane that functions in vacuole fusion and inheritance.

98158703:
Vac8p, a vacuolar protein with armadillo repeats, functions in both vacuole inheritance and protein targeting from the cytoplasm to vacuole.

Peptide information for frame 3

ORF from 99 bp to 1613 bp; peptide length: 505
Category: similarity to known protein
Classification: unset

```

1 MVNILDSPHK SLKCLAAETI ANVAKFKRAR RVVRQHGGIT KLVALLDCAH
51 DSTKPAQSSL YEARDVEVAR CGALALWSCS KSHTNKEAIR KAGGIPLLAR
101 LLKTSHENML IPVVGTLQEC ASEENYRAAI KAERIENLV KNLSSENEQL
151 QEHCAMAIYQ CAEDKETRDL VRLHGGLKPL ASLLNNTDNK ERLAAVTGAI
201 WKCSISKENV TKFREYKAIE TLVGLLTDQP EEVLNVVGA LGECCQEREN
251 RVIVRKC GGI QPLVNLLVGI NQALLVNVTK AVGACAVEPE SMIIIDRLDG
301 VRLWLSLKN PHPDVKASAA WALCPCIKNA KDAGEMVRSF VGGLELIVNL
351 LKSDNKEVLA SVCAAITNIA KDQENLAVIT DHGVVPLLSK LANTNNNKLK
401 HHLAEAISRC CMWGRNRVAF GEHKAVAPLV RYLSNDTNV HRATAQALYQ
451 LSEDADCIT MHENGAVKLL LDMVGSPDQD QEAAAGCIS NIRRLALATE
501 KARYT

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_35p17, frame 3

PIR:S50446 VAC8 protein - yeast (*Saccharomyces cerevisiae*), N = 1,
Score = 237, P = 7.8e-17

PIR:T00403 T13E15.9 protein - *Arabidopsis thaliana*, N = 1, Score = 215,
P = 4.9e-14

TREMBL:DR41081_1 product: "b-catenin"; Danio rerio b-catenin mRNA,
complete cds., N = 1, Score = 195, P = 5.8e-12

>PIR:S50446 VAC8 protein - yeast (*Saccharomyces cerevisiae*)
Length = 578

HSPs:

Score = 237 (35.6 bits), Expect = 7.8e-17, P = 7.8e-17
Identities = 106/401 (26%), Positives = 177/401 (44%)

```

Query:  92 AGGIPLLARLLKTSHENMLIPVVGTLQECASEENYRAAIKAERIENLVKNLSSENEQLQ 151
      +GG PL A      +N+ +      L      E Y  + E ++E ++ L S++ Q+Q
Sbjct:  45 SGG-PLKALTTLVYSDNLNQRSAAALFAEITEKYVRQVSRE-VLEPILILLQSQDPQIQ 102

Query:  152 EHCAMAIYQCAEDKETRDLVRLHGGLKPLASLLNNTDNKERLAAVTGAIWKCSISKENVT 211
      A+  A + E + L+  GGL+PL + +  DN E      G I  +  +N
Sbjct:  103 VAACAALGNLAVNNENKLLIVEMGGLEPLINQMMG-DNVEVQCNAVGCITNLATRDDNKH 161

Query:  212 KFREYKAIE TLVGLLTDQPEEVLNVVVGALGECCQERENRVIVRKC GGIQPLVNLLVGIN 271
      K      A+  L L  +  V N  GAL      ENR +  G +  LV+LL  +
Sbjct:  162 KIATSGALIP LTKLAKSKHIRVQRNATGALLNMTHSENRKELVNAGAVPVLVSLSSSTD 221

Query:  272 QALLVNVTKAVGACAVEPE SMIIIDRLDG--VRLWLSLLKNPHPDVKASAAWALCPCIKN 329
      +      T A+  AV+  +  + + +  V L SL+ +P  VK  A  AL  +

```

Sbjct: 222 PDVQYYCTTALSNIADVDEANRKKLAQTEPRLVSKLVSLMDSPSSRVKQATLALRNLASD 281

Query: 330 AKDAGEMVRSFVGGLELIVNLLKSDNKE-VLASVCAAITNIAKDQENLAVITDHGVV-PL 387
E+VR+ GGL +V L++SD+ VLASV A I NI+ N +I D G + PL

Sbjct: 282 TSYQLEIVRA--GGLPHLVKLIQSDSIPLVLASV-ACIRNISIHLNEGLIVDAGFLKPL 338

Query: 388 LSKLANTNNKLRHHLAEAISRCCMWG-RNRVAFGEHKAVAPLVRYLKSNDTNVHRATAQ 446
+ L ++ +++ H + +NR F E AV + +V ++

Sbjct: 339 VRLLDYKDSEEQCHAVSTLRNLAASSEKNRKEFFESGAVEKCKELALDSPVSV-QSEIS 397

Query: 447 ALYQLESDAD-NCITMHENGAVKLLDMVGSPDQDLQEAAGCISNI 492
A + + AD + + + E + L+ M S +Q++ AA ++N+

Sbjct: 398 ACFAILALADVSKLDLLEANILDALIPMTFSQNEVSGNAAALANL 444

Score = 213 (32.0 bits), Expect = 3.6e-14, P = 3.6e-14
Identities = 81/341 (23%), Positives = 163/341 (47%)

Query: 163 EDKETRDVRLHGGGLKPLASLLNNTD-NKERLAAVTGAIWKCSISKENVTKFREYKAIET 221
EDK+ D G LK L +L+ + + N +R AA+ A I+++ V + + +E

Sbjct: 36 EDKQQLDFYS-GGPLKALTTLVYSDNLNLRSAALAF-----EITEKYVRQVSR-EVLEP 89

Query: 222 LVGLLTDQPEEVLVNVVVGALGECCQERENRVIVRKCGGIQPLVNLVGINQALLVNVTKA 281
++ LL Q ++ V ALG EN++++ + GG++PL+N ++G N + N

Sbjct: 90 ILILQSQDPQIQVAACAALGNLAVNNENKLLIVEMGGLEPLINQMMGDNVEVQCNAVGC 149

Query: 282 VGACAVEPESMMIIDRLDGVRLLSLLKNPHDPVKASAALWALCPCIKNAKDAGEMVRSFV 341
+ A ++ I + L L K+ H V+ +A AL + ++ E+V +

Sbjct: 150 ITNLATRDNDKHKIATSGALIPLTKLAKSKHIRVQRNATGALLNMTHSEENRKELVNA-- 207

Query: 342 GGLELIVNLLKSDNKEVLASVCAAITNIAKDQENLAVI--TDHGVVPLLSKLANTNNK 399
G + ++V+LL S + +V A++NIA D+ N + T+ +V L L ++ +++

Sbjct: 208 GAVPVLVSLSSSTDPDVQYYCTTALSNIADVDEANRKKLAQTEPRLVSKLVSLMDSPSSRV 267

Query: 400 RHHLAEAISRCCMWGRNRVAFGEHKAVAPLVRYLKSNDTNVHRATAQALYQLESDADNCI 459
+ A+ ++ + LV+ ++S+ + A+ + +S N

Sbjct: 268 KCQATLALRNLASDTSYQLEIVRAGGLPHLVKLIQSDSIPLVLASVACIRNISIHLNEG 327

Query: 460 TMHENGAVKLLDMVGSPDQDLQEAAGCISNIRRLALATEKAR 503
+ + G +K L+ ++ D + E +S +R LA ++EK R

Sbjct: 328 LIVDAGFLKPLVRLDYKDSE--EQCHAVSTLRNLAASSEKNR 369

Score = 180 (27.0 bits), Expect = 1.6e-10, P = 1.6e-10
Identities = 80/346 (23%), Positives = 142/346 (41%)

Query: 145 SENEQLEHCAMAIYQCAEDKETRDVRLHGGGLKPLASLLNNTDNKERLAAVTGAIWKCS 204
S+N LQ A+A + E K R + R L+P+ LL + D + ++AA A+ +

Sbjct: 58 SDNLNLRSAALAFAEITE-KYVRQVSR--EVLEPILILQSQDPQIQVAACA-ALGNLA 113

Query: 205 ISKENVTKFREYKAIETLVGLLTDQPEEVLVNVVVGALGECCQERENRVIVRKCGGIQPLV 264
++ EN E +E L+ + EV N VG + +N+ + G + PL

Sbjct: 114 VNNENKLLIVEMGGLEPLINQMMGDNVEVQCNAVGCITNLATRDNDKHKIATSGALIPLT 173

Query: 265 NLLVGINQALLVNVTKAVGACAVEPESMMIIDRLDGVRLLSLLKNPHDPVKASAALWALC 324
L + + N T A+ E+ + V +L SLL + PDV+ AL

Sbjct: 174 KLAKSKHIRVQRNATGALLNMTHSEENRKELVNAGAVPVLVSLSSSTDPDVQYYCTTALS 233

Query: 325 PCIKNAKDAGEMVRSFVGGLELIVNLLKSDNKEVLASVCAAITNIAKDQENLAVITDHGV 384
+ + ++ ++ + +V+L+ S + V A+ N+ A D I G

Sbjct: 234 NIAVDEANRKKLAQTEPRLVSKLVSLMDSPSSRVKQATLALRNLASDTSYQLEIVRAGG 293

Query: 385 VPLLSKLANTNNKLRHHLAEAISRCCMWGRNRVAFGEHKAVAPLVRYLKSNDTNVHRAT 444
+P L KL +++ L I + N + + PLVR L D+ +

Sbjct: 294 LPHLVKLIQSDSIPLVLASVACIRNISIHLNEGLIVDAGFLKPLVRLDYKDSEEQCH 353

Query: 445 A-QALYQLESDAD-NCITMHENGAVKLLDMVGSPDQDLQEAAGCIS 490
A L L+ ++ N E+GAV+ ++ +Q + C +

Sbjct: 354 AVSTLRNLAASSEKNRKEFFESGAVEKCKELALDSPVSVQSEISACFA 401

Score = 155 (23.3 bits), Expect = 8.8e-08, P = 8.8e-08
Identities = 88/401 (21%), Positives = 175/401 (43%)

Query: 60 LYEAR--VEVARCGALALWSCSKSHTNKEAIRKAGGI-PLARLLKTSHENMLIPVVG 116
L +++D ++VA C AL + + ++ NK I + GG+ PL+ +++ + E + VG

Sbjct: 93 LLQSQDPQIQVAACAALG--NLAVNNENKLLIVEMGGLEPLINQMMGDNVE-VQCNAVGC 149

Query: 117 LQECASEENYRAAIKARIENLVKNLNSENEQLEHCAMAIYQCAEDKETR-DLVRLHG 175
+ A+ ++ + I + L K S++ ++Q + A+ +E R +LV G

Sbjct: 150 ITNLATRDNDKHKIATSGALIPLTKLAKSKHIRVQRNATGALLNMTHSEENRKELVNA-G 208

Query: 176 GLKPLASLLNNTDNKERLAAVTGAIWKCSISKENVTKFR--EYKAIETLVGLLTDQPEEV 233
+ L SLL++TD + T A+ ++ + N K E + + LV L+ V

Sbjct: 209 AVPVLVSLSSSTDPDVQYYCTT-ALSNIADVDEANRKKLAQTEPRLVSKLVSLMDSPSSRV 267

Query: 234 LVNVVVGALGECCQERENRVIRKCGGIQPLVNLVGINQALLVNVTKAVGACAVEPESMM 293
AL + ++ + + GG+ LV L+ + L++ + ++ P +

Sbjct: 268 KCQATLALRNLASDTSYQLEIVRAGGLPHLVKLIQSDSIPLVLASVACIRNISHPLNEG 327

Query: 294 IIDRLDGVRLWLSLLK-NPHPDVKASAALWCPCIKNA-KDAGEMVRSFVGGLELIVNLL 351
+I ++ L LL +++ A L ++ K+ E S G +E L

Sbjct: 328 LIVDAGFLKPLVRLLDYKDSEEIQCHAVSTLRNLAASSEKNRKEFFES--GAVEKCKELA 385

Query: 352 KSDNKEVLA--SVCAAITNIAKDQENLAVITDHGVVPLLSKLANTNNKLRHHLAEAIRS 409
V + S C AI +A D L ++ + ++ L + + N ++ + A A++

Sbjct: 386 LDSPVSVQSEISACFAILALA-DVSKDLL-EANILDALIPMTFSQNEVSGNAAAALAN 443

Query: 410 CCMWGRNRVAFGE----HKAVAP-LVRYLKSNDTNVHRATAQALYQLSE 453
C N E ++ + L+R+LKS+ + QL E

Sbjct: 444 LCSRVNNYTKIIEAWDRPNEGIRGFLIRFLKSDYATFEHIALWTILQLE 493

Score = 139 (20.9 bits), Expect = 5.0e-06, P = 5.0e-06
Identities = 80/329 (24%), Positives = 142/329 (43%)

Query: 37 GGITKLVALLDCAHD-STKPAQ---SSLYEARDVEVARCGALALWSCSKSHTNKEAIRKA 92
G I T L D H +T A + L +++ + V R AL + + S N++ + A

Sbjct: 148 GCITNLATRDDNKHKIATSGALIPLTKLAKSKHIRVQRNATGALLNMTHSEENRKELVNA 207

Query: 93 GGIPLLARLLKTSHENMLIPVVGTLQECASEE-NYRAAIKAE-RIENLVKNLNSENEQL 150
G +P+L LL ++ ++ L A +E N + + E R++ LV ++S + ++

Sbjct: 208 GAVPVLVSLSSSTDPDVQYYCTTALSNIADVDEANRKKLAQTEPRLVSKLVSLMDSPSSRV 267

Query: 151 QEHCAAIYQCAEDKETR-DLVRLHGGGLKPLASLLNNTDNKERLAAVTGAIWKCSISKEN 209
+ +A+ A D + ++VR GGL L L+ + D+ + A I SI N

Sbjct: 268 KCQATLALRNLASDTSYQLEIVRA-GGLPHLVKLIQS-DSIPLVLASVACIRNISHPLN 325

Query: 210 VTKFREYKAIETLVGLLT-DQPEEVLVNVVVGALGECCQERE-NRVIVRKCAGGIQPLVNL 267
+ ++ LV LL EE+ + V L E NR + G ++ L

Sbjct: 326 EGLIVDAGFLKPLVRLLDYKDSEEIQCHAVSTLRNLAASSEKNRKEFFESGAVEKCKELA 385

Query: 268 VG--INQALLVNVTKAVGACA-VEPESMMIIDRLDGVRLWLSLLKNPHPDVKASAALA-L 323
+ ++ ++ A+ A A V ++ + LD + + + +N A+AA A L

Sbjct: 386 LDSPVSVQSEISACFAILALADVSKDLLLEANILDAL-IPMTFSQNEVSGNAAAALANL 444

Query: 324 CPCIKN-AKDAGEMVRSFVGGLELIVNLLKSD 354
C + N K R G ++ LKSD

Sbjct: 445 CSRVNNYTKIIEAWDRPNEGIRGFLIRFLKSD 476

Score = 136 (20.4 bits), Expect = 1.1e-05, P = 1.1e-05
Identities = 72/304 (23%), Positives = 133/304 (43%)

Query: 58 SSLYEARDVEVARCGALALWSCSKSHTNKEAIRKAGGIPLLARLLKTSHENMLIPVVGTL 117
+ L +++ + V R AL + + S N++ + AG +P+L LL ++ ++ L

Sbjct: 173 TKLAKSKHIRVQRNATGALLNMTHSEENRKELVNAGAVPVLVSLSSSTDPDVQYYCTTAL 232

Query: 118 QECASEE-NYRAAIKAE-RIENLVKNLNSENEQLQEHCAAIYQCAEDKETR-DLVRLH 174
A +E N + + E R++ LV ++S + +++ +A+ A D + ++VR

Sbjct: 233 SNIADVDEANRKKLAQTEPRLVSKLVSLMDSPSSRVKQATLALRNLASDTSYQLEIVRA- 291

Query: 175 GGLKPLASLLNNTDNKERLAAVTGAIWKCSISKENVTKFREYKAIETLVGLLT-DQPEEV 233
GGL L L+ + D+ + A I SI N + ++ LV LL EE+

Sbjct: 292 GGLPHLVKLIQS-DSIPLVLASVACIRNISHPLNEGLIVDAGFLKPLVRLLDYKDSEEI 350

Query: 234 LVNVVVGALGECCQERE-NRVIVRKCAGGIQPLVNLVVG--INQALLVNVTKAVGACA-VEP 289
+ V L E NR + G ++ L + ++ ++ A+ A A V

Sbjct: 351 QCHAVSTLRNLAASSEKNRKEFFESGAVEKCKELALDPSVSVQSEISACFAILALADVSK 410

Query: 290 ESMMIIDRLDGVRLWLSLLKNPHPDVKASAALA-LCPCIKN-AKDAGEMVRSFVGGLELI 347
++ + LD + + + +N A+AA A LC + N K R G +

Sbjct: 411 LDLEANILDAL-IPMTFSQNEVSGNAAAALANLCSRVNNYTKIIEAWDRPNEGIRGFL 469

Query: 348 VNLLKSD 354
+ LKSD

Sbjct: 470 IRFLKSD 476

Score = 114 (17.1 bits), Expect = 2.7e-03, P = 2.7e-03
Identities = 71/335 (21%), Positives = 132/335 (39%)

Query: 1 MVNILDSPHKSCLKAAETIANVAKFKRARRVVRQHGGITKLVALLDCAHDSTKPAQSSL 60
+ + S H ++ A + N+ + R+ + G + LV+LL ST P

Sbjct: 172 LTKLAKSKHIRVQRNATGALLNMTHSEENRKELVNAGAVPVLVSLSS----STDP----- 222

Query: 61 YEARDVEVARCGALALWSCSKSHTNKEAIRKAGGIPLLARLLKTSHENMLIPVVGTLQEC 120
DV+ AL+ + +++ K A + + L L+ + + L+

Score = 106 (15.9 bits), Expect = 2.0e-02, P = 2.0e-02
Identities = 49/204 (24%), Positives = 89/204 (43%)

Pedant information for DKFZphtes3_35p17, frame 3

Report for DKFZphtes3_35p17.3

```
SEQ      MVNILDSPHSLKCLAAETIANVAKFKRARRVVVRQHGGITKLVALLDCAHDSTKPQSSIL  
SEG      ..... xxxxxxxxxxxxxxxx .....  
2bct-    ..... HH
```

```
SEQ      YEARDVEVARCGALALWSCSKSHTNKEAIRKAGGIPLLARLLKTSHENMLIPVVGTLQECC  
SEG      ....  
2bct-    HHCSCCHHHHHHHHHHHHHHHHHHHHHCCCCHHHHHHHGGGCCCHHHHHHHHHHHHHHHHH
```

```
SEQ      ASEENYRAAIIKAERTIENLVKNLNSENEQLQEHCAMAIYQAEDKETRDVLRLHGGLKPL  
SEG      ....  
2bct-    HNTTTNNHHHHHHCHHHHHHHHHCCCCHHHHHHHHHHHHHHHTTNNHHHHHHHHCHHHHH
```

```
SEQ    ASLLNNTDNKERLAAVTGAIWKCSISKENVTKFREYKAIETLVGLLTDQPEEVLVNVVGA
SEG    .....
2bct-  HHHHH-HCCCHHHHHHHHHHHHHHCCCHHHHHHHHHCHHHHHHTTTTCCHHHHHHHHHHH

SEQ    LGECQERENRVIVRKCGGIQPLVNLVGINQALLVNVTKAVGACAVEPESMMIIDRLDG
SEG    .....
2bct-  H-----HHHHHCCCTTTHHHHHHHHHHHCTTTHHHHHHHHTTTHHHHHHHH-HHCH

SEQ    VRLLSLLKNPHDPVKASAAWALCPCIKNAKDAGEMVRSFVGLELIVNLLKSDNKEVLA
SEG    .....
2bct-  HHHHHHHHTTTHHHHHHHHHHHHHHCCCHH-HHHHHHHHHHHHHHHCTTTTTHHHH

SEQ    SVCAAITNIAKDQENLAVITDHGVVPLLSKLANTNNKLRHHLAEAISRCCMWGRNRVAF
SEG    .....
2bct-  HHHHHHHHHHC GGGHHHHHHCHHHHHHHHHHHHTTTCCHHHHHHHHHHHCHHHHH

SEQ    GEHKAVAPLVRYLKSNDTNVHRATAQALYQLSEADNCITMHENGAVKLLDMVGSPDQD
SEG    .....
2bct-  HTTTHHHHHHHHCCCHHHHHHHHHHHHTTTHHHHHHHHCHHHHHHTTTTTHH

SEQ    LQEAAAGCISNIRRLALATEKARYT
SEG    .....
2bct-  HHHHHHHHH.....
```

(No Prosite data available for DKFZphtes3_35p17.3)

(No Pfam data available for DKFZphtes3_35p17.3)

DKFZphtes3_35p22

group: cell cycle

DKFZphtes3_35p22 encodes a novel 549 amino acid protein, with similarity to oncogene 1 (tre-2 locus).

The novel protein is closely related to human tre-2 and other enzymes involved in the degradation of ubiquitinated proteins. The human tre-2 oncogene encodes a deubiquitinating enzyme, indicating a role for the ubiquitin system in mammalian growth control.

The novel protein can find application in cancer diagnostics and treatment, and in regulating protein stability and growth control via regulation of ubiquitination.

strong similarity to oncogene 1 (tre-2 locus)

membrane regions: 1

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: map="17"

Insert length: 2072 bp

Poly A stretch at pos. 2062, polyadenylation signal at pos. 2039

```
1 GTTACACACA GGCAGTGGTA TCTGTGAGCA GCTCTGTGGA CTCAAAGGTT
51 TTCTCCCTGA GAGGCATGAC CCAGGCCAGC TGATTCATCA GAATCAGGAT
101 GGACGTGGTA GAGGTCGCGG GCAGTTGGTG GGCACAAGAG CGAGAGGACA
151 TCATTATGAA ATACGAAAAG GGACACCGAG CTGGGCTGCC AGAGGACAAG
201 GGGCCTAAGC CTTTTCGAAG CTACAACAAC AACGTCGATC ATTTGGGGAT
251 TGTACATGAG ACGGAGCTGC CTCCTCTGAC TGCAGCGGAG CGGAAGCAAA
301 TTCGGCGGGA GATCAGCCGA AAGAGCAAGT GGGTGGATAT GCTGGGAGAC
351 TGGGAGAAAT ACAAAGCAG CAGAAAGCTC ATAGATCGAG CGTACAAGGG
401 AATGCCCATG AACATCCGGG GCCCGATGTG GTCAGTCCTC CTGAACACTG
451 AGGAAATGAA GTTGAAAAAC CCCGGAAGAT ACCAGATCAT GAAGGAGAAG
501 GGCAGAAGT CATCTGAGCA CATCCAGCGC ATCAGCCGGG ACGTAAGCGG
551 GACATTAAGG AAGCATATAT TCTTCAGGGA TCGATACGGA ACCAAGCAGC
601 GGGAACTACT CCACATCCTC CTGGCATATG AGGAGTACAA CCCGGAGGTG
651 GGCTACTGCA GGGACCTGAG CCACATCGCC GCCTGTGTCC TCCTCTATCT
701 TCCTGAGGAG GATGCATTCT GGGCACTGGT GCAGCTGCTG GCCAGTGAGA
751 GGCATCCCTT GCAGGGATT CACAGCCCAA ATGGCGGGAC CGTCCAGGGG
801 CTCGAAGACC AACAGGAGCA TGTGGTAGCC ACGTCACAAC CCAAGACCAT
851 GGGGCATCAG GACAAGAAAG ATCTATGTGG GCAGTGTTC CCGTTAGGCT
901 GCCTCATCCG GATATTGATT GACGGGATCT CTCTCGGGCT CACCCTGCGC
951 CTGTGGGAGC TGTATCTGGT AGAAGGCGAA CAGGCGCTGA TGCCGATAAC
1001 AAGAATCGCC TTTAAGGTTT AGCAGAAGCG CCTCACGAAG ACGTCCAGGT
1051 GTGGCCCGTG GGCACGTTTT TGCAACCGGT TCGTTGATAC CTGGGCCAGG
1101 GATGAGGACA CTGTGCTCAA GCATCTTAGG GCCTCTATGA AGAAACTAAC
1151 AAGAAAGAAG GGGGACCTGC CACCCCAAGC CAAACCCGAG CAAGGGTCGT
1201 CGGCATCCAG GCCTGTGCGG GCTTCACGTG GCGGGAAGAC CCTCTGCAAG
1251 GGGGACAGGC AGGCCCCCTC AGGCCCCACA GCCCGGTTC CGCGGCCCAT
1301 TTGGTCAGCT TCCCGCCAC GGGCACCTCG TTCTTCCACA CCCTGTCTTG
1351 GTGGGGCTGT CCGGAAGAC ACCTACCTTG TGGGCACTCA GGGTGTGCCC
1401 AGCCCCGGCC TGGCTCAGGG AGGACCTCAG GGTTCCTGGA GATTCTTGCA
1451 GTGGAATCC ATGCCCGGCC TCCCAACGGA CCTGGACGTA GAGGGCCCTT
1501 GGTCCGCCA TTATGATTTC AGACAGAGCT GCTGGGTCCG TGCCATATCC
1551 CAGGAGGACC AGCTGGCCCC CTGCTGGCAG GCTGAACACC CTGCGGAGCG
1601 GGTGAGATCG GCTTTCGCTG CACCCAGCAC TGATTCCGAC CAGGGCACCC
1651 CCTTCAGAGC TAGGGACGAA CAGCAGTGTG CTCACACCTC AGGGCCTTGC
1701 CTCTGCGGCC TCCACTTGA AAGTTCCTAG TTCCCTCCAG GCTTCTAGAA
1751 GCATCTGGGC CAGGGCTCAT GGCTGGATAA TTTCCCTAGG CTTAACAACC
1801 CAAGCAAGCT TCGCATCCTC GTTTTATTTT TGGTAAACT TATGAAAATG
1851 TATTAAGAAA GAGTGCAGCT CGAGAGAGAT TCAGAGATGG AACACACCAG
1901 ACCCCAGATC ACAAAGCCAA CCATGCCAG CCCCTCCAG CACCCCAAGC
1951 CCCACGACCA TCGTCTGTAA TTCTGACGAC ACCGTGAGCC TGCCTTTGTA
2001 CTTCAAACTC ATGGAAGGAT AACCACCTTC ATGTTTGAA ATAAATGTTT
2051 CCTGTTGAAA TGAAAAAAA AA
```

BLAST Results

Entry AC003976 from database EMBL:
Homo sapiens chromosome 17, clone hCIT.91.J.4, complete sequence.
Score = 4385, P = 0.0e+00, identities = 881/886

14 exons

Entry HSG19723 from database EMBL:

human STS A001W35.

Score = 850, P = 1.9e-32, identities = 170/170

Medline entries

92228503:

A novel transcriptional unit of the tre oncogene widely expressed in human cancer cells.

94067315:

The yeast DOA4 gene encodes a deubiquitinating enzyme related to a product of the human tre-2 oncogene.

95176708:

UBP5 encodes a putative yeast ubiquitin-specific protease that is related to the human Tre-2 oncogene product.

Peptide information for frame 3

ORF from 99 bp to 1745 bp; peptide length: 549
Category: strong similarity to known protein

```

1 MDVVEVAGSW WAQEREDIIM KYEKGHRAGL PEDKGPKPFR SYNNNVVDHLG
51 IVHETELPPL TAREAKQIRR EISRKSKWVD MLGDWEKYKS SRKLIDRAYK
101 GMPMNIRGPM WSVLLNTEEM KLKNPGRYQI MKEKGKKSSE HIQRIDRDVS
151 GTLRKHIFFR DRYGTKQREL LHILLAYEY NPEVGYCRDL SHIAALFLLY
201 LPEEDAFWAL VQLLASERHS LQGFHSPNGG TVQGLQDQQE HVVATSQPKT
251 MGHQDKKDLG GQCSPLGCLI RILIDGISLG LTLRLWDVYL VEGEQALMPI
301 TRIAFKVQQK RLTKTSRCGP WARFCNRFVD TWARDLTVL KHLRASMKKL
351 TRKKGDLPPP AKPEQGSSAS RVPASRGGK TLCKGDRQAP PGPPARFPPR
401 IWSASPPRAP RSSTPCPGA VREDTYPVGT QGVSPALAQ GGPQGSWRFL
451 QWNSMPRLPT DLDVEGPWFR HYDFRQSCWV RAISQEDQLA PCWQAEHPAE
501 RVRSAFAAPS TDSQGTFFR ARDEQQCAPT SGFCLCGLHL ESSQFPFPGF

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_35p22, frame 3

PIR:S22155 oncogene 1 (tre-2 locus) (clone 210) - human, N = 1, Score = 2181, P = 5.5e-226

PIR:S57867 oncogene 1 - human, N = 1, Score = 1536, P = 1.2e-157

>PIR:S22155 oncogene 1 (tre-2 locus) (clone 210) - human
Length = 786

HSPs:

Score = 2181 (327.2 bits), Expect = 5.5e-226, P = 5.5e-226
Identities = 405/500 (81%), Positives = 440/500 (88%)

```

Query:      1 MDVVEVAGSWWAQEREDIIMKYEKGHRAGLPEDKGPKPFRSYNNNVVDHLGIVHETELPPL 60
            MD+VE A S AQER+DI+MKY+KGHRAGLPEDKG+P N+++D GI+HETELPP+
Sbjct:      1 MDMVENADSLQAQERKDILMKYDKGHRAGLPEDKGPEPV-GINSSIDRFGLHETELPPV 59

Query:      61 TAREAKQIRREISRKSKWVDMLGDWEKYKSSRKLIDRAYKGMPMNIRGPMWSVLLNTEEM 120
            TAREAK+IRRE++R SKW++MLG+WE YK S KLIDR YKG+PMNIRGP+WSVLLN +E+
Sbjct:      60 TAREAKKIRREMTRTSKWMEMLGEWETYKHSSKLIDRVYKGIPIRNPVWSVLLNIQEI 119

Query:      121 KLKNPGRYQIMKEKGKKSSEHIQRIDRDVSGTLRKHIFFRDRYGTKQRELHILLAYEY 180
            KLKNPGRYQIMKE+GK+SSEHI ID DV TLR H+FFRDRYG KQREL +ILLAY EY
Sbjct:      120 KLKNPGRYQIMKERGRSSEHIHHIDLVRTTLRNHVFFRDRYGAKQRELFYILLAYSEY 179

Query:      181 NPEVGYCRDLSHIAALFLLYLPEEDAFWALVQLLASERHSLQGFHSPNGGTVQGLQDQQE 240
            NPEVGYCRDLSHI ALFLLYLPEEDAFWALVQLLASERHSL GFHSPNGGTVQGLQDQQE
Sbjct:      180 NPEVGYCRDLSHITALFLLYLPEEDAFWALVQLLASERHSLPGFHSPPNGGTVQGLQDQQE 239

```

Query: 241 HVVATSQPKTMGHQDKKDLGGCQSPGLCLIRILIDGISLGLTLRLWDVYLVEGEQALMPI 300
 HVV SQPKTM HQDK+ LCGQC+ LGCL+R LIDGISLGLTLRLWDVYLVEGEQ LMPI
 Sbjct: 240 HVVPKSQPKTMHGHQDKKDLGGCQSPGLCLIRILIDGISLGLTLRLWDVYLVEGEQVLMPI 299

Query: 301 TRIAFKVQKRLTKTSRCGPWARFCNRFVDTWARDEDTVLKHLRASMKKLTRKKGDLPPP 360
 T IA KVQKRL KTSRCG WAR N+F DTWA ++DTVLKHLRAS KKLTRK+GDLPPP
 Sbjct: 300 TSIALKVQKRLMKTSRCGLWARLRNQFFDTWAMNDDTVLKHLRASTKKLTRKQGDLP 359

Query: 361 AKPEQGSSASRPVPASRGKTLCKGDRQAPPGPAPRFRPIWSASPPRAPSSTPCPGGA 420
 AK EQGS A RVPVPASRGKTLCKG RQAPPGPPA+F RPI SASPP A R STPCPGGA
 Sbjct: 360 AKRQGS LAPRPVPASRGKTLCKGYRQAPPGPPAQFORPICSASPPWASRSTPCPGGA 419

Query: 421 VREDTYPVGTQGVPSPALAQGGPQGSWRFLQWNSMPRLPTDLDVEGPWFRHYDFRQSCWV 480
 VREDTYPVGTQGVPS ALAQGGPQGSWRFL+W SMPRLPTDLD+ GPWF HYDF +SCWV
 Sbjct: 420 VREDTYPVGTQGVPSLALAQGGPQGSWRFLWKSMPLPTDLDIGGPWFPHYDFERSCWV 479

Query: 481 RAISQEDQLAPCWQAHPAE 500
 RAISQEDQLA CWQAEP E
 Sbjct: 480 RAISQEDQLATCWQAHPCE 499

Pedant information for DKFZphtes3_35p22, frame 3

Report for DKFZphtes3_35p22.3

[LENGTH] 549
 [MW] 62159.16
 [pI] 9.23
 [HOMOL] PIR:S22155 oncogene 1 (tre-2 locus) (clone 210) - human 0.0
 [FUNCAT] 11.01 stress response [S. cerevisiae, YGR100w] 2e-16
 [FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YGR100w] 2e-16
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YNL293w] 3e-15
 [PIRKW] transmembrane protein 6e-14
 [PROSITE] MYRISTYL 6
 [PROSITE] AMIDATION 1
 [PROSITE] CAMP_PHOSPHO_SITE 3
 [PROSITE] CK2_PHOSPHO_SITE 4
 [PROSITE] TYR_PHOSPHO_SITE 2
 [PROSITE] PKC_PHOSPHO_SITE 10
 [KW] TRANSMEMBRANE 1
 [KW] LOW_COMPLEXITY 5.28 %

SEQ MDVVEVAGSWWAQEREDIIMKYEKGHRAGLPEDKGPKPFRSYNNNVNDHLGIVHETELPPL
 SEG
 PRD cccceeeccchhhhhhhhhhhhhhhccccccccccccccccceeecccccccccccccccc
 MEM

SEQ TAREAKQIRREISRSKSWDMLGDWEKYSSRKLIIDRAYKGMPMNIRGPMWSVLLNTEEM
 SEG
 PRD chhhhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhccccccccceeecccccc
 MEM

SEQ KLKNPGRYQIMKEKGKKSSEHIQRIDRDVSGTLRKHIFFRDRYGTQRELLHILLAYEY
 SEG
 PRD cccccccchhhhhhhhhccccchhhhhhhhhhhhhccccccccccccccccchhhhhhhhhhhc
 MEM

SEQ NPEVGYCRDLSHIAALFLLYLPEEDAFWALVQLLASERHSIQGFHSPNGGTVQGLDQQE
 SEG
 PRD cccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccchhhhhhhhh
 MEM

SEQ HVVATSQPKTMGHQDKKDLGGCQSPGLCLIRILIDGISLGLTLRLWDVYLVEGEQALMPI
 SEG
 PRD hhhhhhhchhhhhhhhhccccccccchhhhhhhhhhhccccchhhhhhhhhccccceeehh
 MEMMMMMMMMMMMMMMMMM.....

SEQ TRIAFKVQKRLTKTSRCGPWARFCNRFVDTWARDEDTVLKHLRASMKKLTRKKGDLPPP
 SEG
 PRD hhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhccccc
 MEM

SEQ AKPEQGSSASRPVPASRGKTLCKGDRQAPPGPAPRFRPIWSASPPRAPSSTPCPGGA
 SEG
 PRD ccc
 MEM

```

SEQ      VREDTYPVGTQGVSPALAQGGPQGSWRFLQWNSMPRLPTDLDEGPWFRHYDFRQSCWV
SEG      .....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM      .....

SEQ      RAISQEDQLAPCWQAEHPAERVRSFAAPSTDSDQGTFFRARDEQQCAPTSGPCLCGLHL
SEG      .....
PRD      cchhhhhhhhhhhhhhhcchhhhhhhhhccccccccccccchhhhhccccccccceeee
MEM      .....

SEQ      ESSQFPFPGF
SEG      .....
PRD      ccccccccc
MEM      .....

```

Prosites for DKFZphtes3_35p22.3

PS00004	136->140	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	310->314	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	348->352	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	61->64	PKC_PHOSPHO_SITE	PDOC00005
PS00005	73->76	PKC_PHOSPHO_SITE	PDOC00005
PS00005	90->93	PKC_PHOSPHO_SITE	PDOC00005
PS00005	152->155	PKC_PHOSPHO_SITE	PDOC00005
PS00005	216->219	PKC_PHOSPHO_SITE	PDOC00005
PS00005	282->285	PKC_PHOSPHO_SITE	PDOC00005
PS00005	315->318	PKC_PHOSPHO_SITE	PDOC00005
PS00005	346->349	PKC_PHOSPHO_SITE	PDOC00005
PS00005	351->354	PKC_PHOSPHO_SITE	PDOC00005
PS00005	446->449	PKC_PHOSPHO_SITE	PDOC00005
PS00006	61->65	CK2_PHOSPHO_SITE	PDOC00006
PS00006	460->464	CK2_PHOSPHO_SITE	PDOC00006
PS00006	484->488	CK2_PHOSPHO_SITE	PDOC00006
PS00006	511->515	CK2_PHOSPHO_SITE	PDOC00006
PS00007	93->100	TYR_PHOSPHO_SITE	PDOC00007
PS00007	92->100	TYR_PHOSPHO_SITE	PDOC00007
PS00008	8->14	MYRISTYL	PDOC00008
PS00008	101->107	MYRISTYL	PDOC00008
PS00008	230->236	MYRISTYL	PDOC00008
PS00008	276->282	MYRISTYL	PDOC00008
PS00008	366->372	MYRISTYL	PDOC00008
PS00008	441->447	MYRISTYL	PDOC00008
PS00009	134->138	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3_35p22.3)

DKFZphtes3_4b4

group: testes derived

DKFZphtes3_4b4 encodes a novel 497 amino acid protein similar to SCP proteins and a human trypsin inhibitor.

The novel protein contains an extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 signature 2, predicted by Prosite and Pfam. This domain is found in a variety of extracellular proteins from eukaryotes that have been found to be evolutionary related. The exact function of these proteins is not yet known. In addition, the protein is similar to a human trypsin inhibitor.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes or as a new protease inhibitor.

strong similarity to trypsin inhibitor

might be a new protease inhibitor?

Sequenced by AGOWA

Locus: /map="333.4 cR from top of Chr16 linkage group"

Insert length: 4574 bp

Poly A stretch at pos. 4551, polyadenylation signal at pos. 4539

```
1  GCGGGCTGCT CCCATTGAGC TGTCTGCTCG CTGTGCCCGC TGTGCCCTGCT
51  GTGCCCCGCG TGTCGCCGCT GCTACCGCGT CTGCTGGACG CGGGAGACGC
101 CAGCGAGCTG GTGATTGGAG CCCTGCGGAG AGCTCAAGCG CCCAGCTCTG
151 CCCGAGGAGC CCAGGCTGCC CCGTGAGTCC CATAGTTGCT GCAGGAGTGG
201 AGCCATGAGC TGCGTCCTGG GTGGTGTGTC CCCCTGGGGG CTGCTGTTCC
251 TGGTCTGCGG ATCCCAAGGC TACCTCCTGC CCAACGTGTC TCTCTTAGAG
301 GAGCTGCTCA GCAATACCA GCACAACGAG TCTCACTCCC GGGTCCGCAG
351 AGCCATCCCC AGGGAGGACA AGGAGGAGAT CCTCATGCTG CACAACAAGC
401 TTCGGGGCCA GGTGCAGCCT CAGGCCTCCA ACATGGAGTA CATGACCTGG
451 GATGACGAAC TGGAGAAGTC TGCTGCAGCG TGGGCCAGTC AGTGCATCTG
501 GGAGCACGGG CCCACCAATC TGCTGGTGTG CATCGGGCAG AACCTGGGCG
551 CTCCTGCGGG CAGGTATCGC TCTCCGGGGT TCCATGTGCA GTCCTGGTAT
601 GACGAGGTGA AGGACTACAC CTACCCCTAC CCGAGCGAGT GCAACCCCTG
651 GTGTCAGAGG AGGTGCTCGG GGCCTATGTG CACGCACTAC ACACAGATAG
701 TTTGGGCCAC CACCAACAAG ATCGGTTGTG CTGTGAACAC CTGCCGGAAG
751 ATGACTGTCT GGGGAGAAGT TTGGGAGAAC GCGGTCTACT TTGTCTGCAA
801 TTATTCTCCA AAGGGGAAGT GGATTGGAGA AGCCCCCTAC AAGAATGGCC
851 GGCCTGCTCT TGAGTGCCCA CCCAGCTATG GAGGCAGCTG CAGGAACAAC
901 TTGTGTTACC GAGAAGAAGC CTACACTCCA AAACCTGAAA CGGACGAGAT
951 GAATGAGGTG GAAACGGCTC CCATTCTGTA AGAAAACCAT GTTTGGCTCC
1001 AACCGAGGGT GATGAGACCC ACCAAGCCCA AGAAAACCTC TGCGGTCAAC
1051 TACATGACCC AAGTCGTCAG ATGTGACACC AAGATGAAGG ACAGGTGCAA
1101 AGGTGCCACG TGTAACAGGT ACCAGTGCCC AGCAGGCTGC CTGAACCACA
1151 AGGCGAAGAT CTTTGGAAGT CTGTTCTATG AAAGCTCGTC TAGCATATGC
1201 CGCGCCGCCA TCCACTACGG GATCCTGGAT GACAAGGGAG GCCTGGTGGA
1251 TATCACCAGG AACGGGAAGG TCCCTTCTT CGTGAAGTCT GAGAGACACG
1301 GCGTGCAAGT CCTCAGCAAA TACAAACCTT CCAGCTCATT CATGGTGTCA
1351 AAAGTGAAAG TGCAGGATTG GACTGCTTAC ACGACCGTTG CTCAGCTGTG
1401 CCCGTTTGAA AAGCCAGCAA CTCCTGCCCC AAGAATCCAT TGTCGGGCAC
1451 ACTGCAAGA CGAACCTTCC TACTGGGCTC CGGTGTTTGG AACCAACATC
1501 TATGACAGTA CCTCAAGCAT CTGCAAGACA GCCGTGCACG CGGGAGTCAT
1551 CAGCAACGAG AGTGGGGGTG ACGTGGACGT GATGCCCCGT GATAAAAAGA
1601 AGACCTACGT GGGCTCGCTC AGGAATGGAG TTCAGTCTGA AAGCCTGGGG
1651 ACTCCTCGGG ATGGAAAGGC CTTCCGGATC TTTGCTGTCA GGCAGTGAAT
1701 TTCCAGCACC AGGGGAGAAG GGGCGTCTTC AGGAGGGCTT CGGGGTTTTG
1751 CTTTTATTTT TATTTTGTCA TTGCGGGGTA TATGGAGAGT CAGGAAACTT
1801 CCTTTGACTG ATGTTCAAGT TCCATCACTT TGTGGCCTGT GGGTGAGGTG
1851 ACATGCTATC CCCTCACTGA AGCAACAGCA TCCCAAGGTG CTCAGCCGGA
1901 CTCCTCGGTG CCTGATCCTG CTGGGGCCCG GGGGTCTCCA TCTGGACGTC
1951 CTCCTCTCCT TAGAGATCTG AGCTGTCTCT TAAAGGGGAC AGTTGCCCAA
2001 AATGTTCTCT GCTATGTGTT CTTCTGTTGG TGGAGGAAGT TGATTTCAAC
2051 CTCCTTGCCA AAAGAACAAA CCATTGGAAG CTCACAATTG TGAAGCATTC
2101 ACGGCGTCGG AAGAGGCCCT TTGAGCAAGC GCCAATGAGT TTCAGGAATG
2151 AAGTAGAAGG TAGTTATTTA AAAATAAAAA ACACAGTCCG TCCCTACCAA
2201 TAGAGGAAAA TGTTTAAAT GTTTGCTGGT CAGACAGACA AATGGGCTAG
2251 AGTAAGAGGG CTGCGGGTAT GAGAGACCCC GGCTCCGCCC TGGCACGTGT
2301 CCTTGCTGGC GGCCCGCCAC AGGCCCCCTT CAATGGCCGC ATTCAGGATG
2351 GCTCTATACA CAGCAGTGCT GGTTTATGTA GAGTTCAGCA GTCACCTCAG
2401 AGATGTATCT TGTCTTTGTC AGGCCCTTCA TCTTCATGGC CCACCTGTTT
2451 TCTGCCGTGA CCTTTGGTCC CATTGAGGAC TAAGGATCGG GACCCCTTCT
```



```

2501 TTACCCCTTA CCCATTGTGG CTCCCACCTT GCCTCGGACT GGTTCACGTG
2551 TCCTGGTTCA CACCCAGGAC TTTTCTTTGC AAGCGAACCT GTTTGAAGCC
2601 CAAGTCTTAA CTCCTGGTCT CGTAAGGTTC CACTGAGACG AGATGTCTGA
2651 GAACAACCAA AGAAGGCCTG CTCTTTGCTG CTTTAAAAA ATGACAATTA
2701 AATGTGCAGA TTCCCCACGC ACCCGATGAC CTATTTTTC AGCCGTGGGA
2751 GGAATGGAGT CTTTGGTACA TTCCCTACCG AGGTTAGCAG CTCAGTTTGT
2801 GGTATGAAA CCGTCTGTGG CCTCATGACA GCGAGAGATG GGAATACACT
2851 AGAAGGATCT CTTTCTCTGT TTTTCGTGAA CGACTCTTGC CAAACGTTCC
2901 CGAGGCGCCA AGGAGTGTAG TACACCCTGG CTGCCATCAC TCTATAAAAG
2951 TGCTTCATGA GCCCAGACCA AAAGCCACA GTGAAATGAA GTACCCTTTT
3001 GTAAATAGCA TTTTCTTGA GAAGGTGAAA ATTCCACTCT CTACCACCGG
3051 GCCAGCCAAT AGATCACTTT GGTGAATGCT AGTTTCAAAT TTGATTCAAA
3101 ATATTTCTTA GGTGAAAGAA CTAGCAGAAA GTCAAAACT AAGATACTGT
3151 AGACTGGACA AGAAATTCTA CCTGGGCACC TAGGTGATGC CTTCTTTCTT
3201 TGATTGCCTT TCTAATAAAT GCAGAATCTG AAGGTAAATA GGTTTAAAC
3251 AAAACAAAAA CCCACCCCTT TAAGGAGTTG GTAAAAAGCA GTTCAACTCT
3301 TAGCTTGACT GAGCTAAAAT TCACAGGACT ACGTGCTTTG TGCATTGTAG
3351 TCTAGTCGTA ATTCAATAGT ACTGACTCCT CAGCCCCAAA TGTCGGAGAG
3401 GAAGAAATTC GTCAGCCTGT CAGGTCGTGA GTCCAGTTAC CACCAAACAT
3451 CTGGGAAACT TCTGGGTGCT GGGTGCTCTG CTGCTGGACT TTTGTGGCTG
3501 TGTCTGTGTC TGCAAGATAA ATTAGATCGC CCTGTGGGT TTGCAGAATT
3551 AGTGAAGGGT CCAGGACGAT CCCAGTGGGC TCGCTTCCAA AGCATCCAC
3601 TCAAGGGAGA CTTGAAACTT CCAGTGTGAG TTGACCCCAT CATTTAAAAA
3651 TAAAGTCCCC GGGTTCCTTA ATGCTCTCTT CACTGGGCCT TCCTAGCAGG
3701 ATAGAAAGTC CTTGCCCAGA GCAGGACCTG GCTGTCTTTT TTTTTTTTTT
3751 TTTCCCGAGA CCAAGTTTCA CTCTGTTGCC CAAGGTAGAG TGCAAGTGGC
3801 TGATCTCTGC TCATTGCAAC TGCCGCCTCC CGGGTTCAAG CAATTCTCAT
3851 GCATCAGCCT CCCAAGTACC TGGGACTACA GCGTGAGCT ACCATGCCCC
3901 GCTAATTTTT GTATTTTATG TAGAGATGGG GTTTCATTAT GTTGGCCAGG
3951 CTGGTCTCGA ACTCCTTACC TCAGGTGATC CACCCACCTT GGCCTCCCGA
4001 AGTGTGGGA TTACAGGCAT GAGCCACTGC GCCCGGCAT GGACCTGGCT
4051 GTCTTTATCA TCCCCACAAA CATTTTGAAA CTGGAATATT TGTCTTCAGA
4101 AAATGGAAAC AAGACTATAA ATGATAAGCC CTGTCCCTAG CACCACCTCT
4151 CCTGTGTGTG GAATAGAGGC CCCTCGTCTG ACCAACACTT ACCCTGTGTT
4201 TAAAAAGATC TTGTACCAAG CCAACGGCGT TCCTGGCTCT CCTGCCCACA
4251 GGATGAACAT TTTCCGCTTC CTTAGGAGTT TTGCCCTACC GTATTCCAAA
4301 GCGTGTGCTG GTTCTCTATA TTGTCTGTAG GCTCACTCAG CCCGAGTTT
4351 ATGTGTGTGC TTTTCTCTAT GAAAAATGAT GTATTTTGCT ACTTCCTGTG
4401 TACAAAGTTT TATTGTAAAT GTTTTGTG CTTTGCATGA ACAGGGGCCA
4451 CGTTGTGCA ATTGTTTCAG TAGAACTGGT TTGATTCTA AAATGTTCCT
4501 GTAACATATC TTTTATGAAC AAATCTGAAC AATTTGTGAA ATAAACATT
4551 GAAACCAAAA AAAAAAAAAA AAAA

```

BLAST Results

Entry HS834352 from database EMBL:
human STS WI-15502.
Score = 1331, P = 5.4e-54, identities = 287/301

Medline entries

98146272:
cDNA cloning of a novel trypsin inhibitor with similarity to
pathogenesis-related proteins, and its
frequent expression in human brain cancer cells.

Peptide information for frame 1

ORF from 205 bp to 1695 bp; peptide length: 497
Category: strong similarity to known protein

```

1 MSCVLGGVIP LGLLFLVCGS QGYLLPNVTL LEELLSKYQH NESHRSVRRA
51 IPREDKEEIL MLHNKLRGQV QPQASNMEYM TWDDLEKSA AAWASQCIWE
101 HGPTSLVSI GQNLGAHWGR YRSPGFHVQS WYDEVKDYTY PYPSECNPCW
151 PERCSGPMCT HYTQIVWATT NKIGCAVNTC RKMTVWGEVW ENAVYFVCNY
201 SPKGNWIGEA PYKNGRPCSE CPPSYGGSCR NNLCYREETY TPKPETDEMN
251 EVETAPIPEE NHVWLQPRVM RPTKPKKTSV VNYMTQVVRC DTKMKDRCKG
301 STCNRYQCPA GCLNHKAKIF GTLFYESSSS ICRAAIHYGT LDDKGGGLVDI
351 TRNGKVPFFV KSERHGVQSL SKYKPSSEFM VSKVKVQDLD CYTTVAQLCP
401 FEKPATHCPR IHCPAHCKDE PSYWAPVFGT NIYADTSSIC KTAHVHAGVIS

```

451 NESGGDQDVM PVDKKKTYVG SLRNGVQSES LGTPRDGKAF RIFAVRQ

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_4b4, frame 1

TREMBLNEW:AF109674_1 gene: "Lg11"; product: "late gestation lung protein 1"; Rattus norvegicus late gestation lung protein 1 (Lg11) mRNA, complete cds., N = 1, Score = 968, P = 1.9e-97

TREMBL:D45027_1 product: "25 kDa trypsin inhibitor"; Homo sapiens mRNA for 25 kDa trypsin inhibitor, complete cds., N = 1, Score = 738, P = 4.5e-73

TREMBL:AB009609_1 gene: "HrTT-1"; Halocynthia roretzi HrTT-1 mRNA, complete cds., N = 1, Score = 345, P = 2e-31

PIR:JC5308 testis-specific, vespid, and pathogenesis-related protein 1 precursor - human, N = 1, Score = 337, P = 1.7e-30

>TREMBLNEW:AF109674_1 gene: "Lg11"; product: "late gestation lung protein 1"; Rattus norvegicus late gestation lung protein 1 (Lg11) mRNA, complete cds.

Length = 188

HSPs:

Score = 968 (145.2 bits), Expect = 1.9e-97, P = 1.9e-97
Identities = 160/185 (86%), Positives = 170/185 (91%)

Query: 61 MLHNKLRGQVQPOASNMEYMTWDDLEKSAAAWASQCIWEHGPTSLVLSIGQNLGAHWGR 120
MLHNKLRGQV P ASNMEYMTWD+ELE+SAAAWA +C+WEHGP SLLVLSIGQNL HWGR
Sbjct: 1 MLHNKLRGQVYPASNMEYMTWDEELERSAAAWAQRCLWEHGPASLLVLSIGQNLAVHWGR 60

Query: 121 YRSPGFHVQSWYDEVKDYTYPPSECNPWCPERCSGPMCTHYTQIVWATTNKIGCAVNTC 180
YRSPGFHVQSWYDEVKDYTYPP ECNPWCPCSCG MCTHYTQ+VWATTNKIGCAV+TC
Sbjct: 61 YRSPGFHVQSWYDEVKDYTYPPHECNPWCPERCSGAMCTHYTQMVWATTNKIGCAVHTC 120

Query: 181 RKMTVWGEVWENAVYFVCNYSKGNWIGEAPYKNGRPCSECPSPSYGGSCRNNLCYREETY 240
R M+VWG++WENAVY VCNYSKGNWIGEAPYK+GRPCSECP SYGG CRNNLCYREE Y
Sbjct: 121 RMSVWGDWENAVYLVNYSKGNWIGEAPYKHGRPCSECPSSYGGGCRNNLCYREEHY 180

Query: 241 TPKPE 245
KPE

Sbjct: 181 HQKPE 185

Pedant information for DKFZphtes3_4b4, frame 1

Report for DKFZphtes3_4b4.1

[LENGTH] 497
[MW] 55920.00
[pI] 8.36
[HOMOL] TREMBL:D45027_1 product: "25 kDa trypsin inhibitor"; Homo sapiens mRNA for 25 kDa trypsin inhibitor, complete cds. 6e-78
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YJL078c] 8e-12
[BLOCKS] BL01009E Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 proteins
[BLOCKS] BL01009D Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 proteins
[BLOCKS] BL01009C Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 proteins
[BLOCKS] BL01009A Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 proteins
[PIRKW] glycoprotein 5e-22
[PIRKW] blocked amino end 5e-13
[PIRKW] brain 9e-30
[PIRKW] hydrolase 4e-09
[PIRKW] hemolymph coagulation 4e-09
[PIRKW] zymogen 4e-09
[PIRKW] alternative splicing 4e-09
[PIRKW] sperm 5e-22
[PIRKW] viroid-induced protein 2e-11
[PIRKW] venom 6e-18
[PIRKW] pyroglutamic acid 2e-11
[PIRKW] transmembrane protein 2e-10
[PIRKW] serine proteinase 4e-09
[SUPFAM] C-type lectin homology 4e-09
[SUPFAM] trypsin homology 4e-09

[SUPFAM] complement factor H repeat homology 4e-09
 [SUPFAM] cysteine-rich secretory protein 1 6e-24
 [SUPFAM] pathogenesis-related leaf protein 7e-15
 [PROSITE] MYRISTYL 8
 [PROSITE] CAMP_PHOSPHO_SITE 3
 [PROSITE] CK2_PHOSPHO_SITE 6
 [PROSITE] TYR_PHOSPHO_SITE 1
 [PROSITE] PKC_PHOSPHO_SITE 8
 [PROSITE] ASN_GLYCOSYLATION 3
 [PROSITE] SCP_AG5_PR1_SC7_2 1
 [PFAM] SCP-like extracellular Proteins
 [KW] All Beta
 [KW] SIGNAL_PEPTIDE 23
 [KW] LOW_COMPLEXITY 1.21 %

SEQ MSCVLGGVPLGLLFLVCGSQGYLLPNVTLLLELLSKYQHNEHSRVRAIPREDKEEIL
 SEGxxxxxx.....
 PRD cccccccccccccccccccccchhhhhhhhhhhhhcccccchhhhhhhccchhhhh

SEQ MLHNKLRGQVQPQASNMEYMTWDELEKSAAAWASQCIWEHGPTSLLSIGQNLGAHWGR
 SEG
 PRD hhhhhhhccccccccchhhhhhhhhhhhhhhhhhhhhcccccchhhhhhhccchhhhh

SEQ YRSPGFHVQSWYDEVKDYTYYPSECNPWCPCSGPMCTHYTQIVWATTNKIGCAVNTC
 SEG
 PRD cccccchhhhhhhhhhhcccccchhhhhhhhhhhhhhhhhhhhhcccccchhhhhhhccchhhhh

SEQ RKMTVWGEVWENAVYFVCNYSFKGNWIGEAPYKNGRPCSECPSPSYGGSCRNNLCYREETY
 SEG
 PRD ccc

SEQ TPKPETDEMNEVETAPIPEENHVWLQPRVMRPTKPKKTSAVNYMTQVVRCDTKMKDRCKG
 SEG
 PRD ccc

SEQ STCNRYQCPAGCLNHKAKIFGTLFYESSSSICRAAIHYGILDDKGGGLDITRNGKVPFFV
 SEG
 PRD ccc

SEQ KSERHGVQSLSKYKPSSEFMVSKVKVQDLDCYTTVAQLCPFEPATHCPRIHCPAHCKDE
 SEG
 PRD ecc

SEQ PSYWAPVFGTNIYADTSSICKTAVHAGVISNESGGDVPDVKKKTYVGSRLNGVQSES
 SEG
 PRD cccecececececc

SEQ LGTPRDGKAFRIFAVRQ
 SEG
 PRD ccccccccccccccccc

Prosite for DKFZphtes3_4b4.1

PS00001	27->31	ASN_GLYCOSYLATION	PDOC00001
PS00001	41->45	ASN_GLYCOSYLATION	PDOC00001
PS00001	451->455	ASN_GLYCOSYLATION	PDOC00001
PS00004	181->185	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	276->280	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	464->468	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	170->173	PKC_PHOSPHO_SITE	PDOC00005
PS00005	179->182	PKC_PHOSPHO_SITE	PDOC00005
PS00005	201->204	PKC_PHOSPHO_SITE	PDOC00005
PS00005	228->231	PKC_PHOSPHO_SITE	PDOC00005
PS00005	241->244	PKC_PHOSPHO_SITE	PDOC00005
PS00005	362->365	PKC_PHOSPHO_SITE	PDOC00005
PS00005	471->474	PKC_PHOSPHO_SITE	PDOC00005
PS00005	483->486	PKC_PHOSPHO_SITE	PDOC00005
PS00006	29->33	CK2_PHOSPHO_SITE	PDOC00006
PS00006	75->79	CK2_PHOSPHO_SITE	PDOC00006
PS00006	81->85	CK2_PHOSPHO_SITE	PDOC00006
PS00006	130->134	CK2_PHOSPHO_SITE	PDOC00006
PS00006	453->457	CK2_PHOSPHO_SITE	PDOC00006
PS00006	483->487	CK2_PHOSPHO_SITE	PDOC00006
PS00007	385->393	TYR_PHOSPHO_SITE	PDOC00007
PS00008	111->117	MYRISTYL	PDOC00008
PS00008	115->121	MYRISTYL	PDOC00008
PS00008	174->180	MYRISTYL	PDOC00008
PS00008	204->210	MYRISTYL	PDOC00008

PS00008	227->233	MYRISTYL	PDOC00008
PS00008	300->306	MYRISTYL	PDOC00008
PS00008	447->453	MYRISTYL	PDOC00008
PS00008	470->476	MYRISTYL	PDOC00008
PS01010	195->207	SCP_AG5_PR1_SC7_2	PDOC00772

Pfam for DKFZphtes3_4b4.1

HMM_NAME	SCP-like extracellular Proteins		
HMM	*PQDEQDEWLNkHNDFRQQVGRGLETRGNPGPQPPAsNMnPMVWNDELAt		
	P + ++E+L HN +R QV	P ASNM M+W+DEL +	
Query	52 PREDKEEILMLHNKLRGQVQ-----PQASNMEYMTWDELEK		88
HMM	IAQnWANQCiFDHHDCWNHsnYPYQONIAWWSsTANnPWnWssMIQMWY		
	A WA+QCI +H ++ + S GQN+ + + +++++ +Q+WY		
Query	89 SAAAWASQCIWEHGPTSLVSI---GQNLGAHWG---RYRSPGFHVQSWY		132
HMM	NEvkDYNNYNWNTCKGG.....NNFmVCGHYTQMVWRnTfrIGCGRYICYC		
	+EVKDY Y + + +C HYTQ+VW+ T +IGC+ C+		
Query	133 DEVKDYTYPPYSECNPWCPCRCSGPMCTHYTQIVWATTNKIGCAVNTCRK		182
HMM	NNNWrkPDPWKkhkYVVCNYCPpGNYmN*		
	+ W + W+ +Y VCN Y P+GN+++		
Query	183 MTVW--GEVWENAVYFVCNYSKGNWIG		208

DKFZphtes3_4f17

group: testes derived

DKFZphtes3_4f17 encodes a novel 656 amino acid protein with weak similarity to methyl-CpG-binding proteins.

Methylation at the DNA sequence 5'-CpG is required for mammalian development. Methyl-CpG-binding proteins bind specifically to methylated DNA via a related amino acid motif and can repress transcription. The novel protein does not contain such a motif.
No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to methyl-CpG-binding protein

extension of HS557771/HSZ78337,
there are some differences to these sequences

Sequenced by AGOWA

Locus: /map="18"

Insert length: 2320 bp

Poly A stretch at pos. 2266, polyadenylation signal at pos. 2251

```
1 GGCAGGTTTCG CGGGTCGCTG GCGGGGGTTCG TGAGGGAGTG CGCCGGGAGC
51 GGAGATATGG AGGGAGATGG TTCAGACCCA GAGCCTCCAG ATGCCGGGGA
101 GGACAGCAAG TCCGAGAATG GGGAGAATGC GCCCATCTAC TGCATCTGCC
151 GCAAACCGGA CATCAACTGC TTCATGATCG GGTGTGACAA CTGCAATGAG
201 TGGTTCCATG GGGACTGCAT CCGGATCACT GAGAAGATGG CCAAGGCCAT
251 CCGGGAGTGG TACTGTCTGG AGTGCAGAGA GAAAGACCCC AAGCTAGAGA
301 TTGCTCGGGG GCACAAGAAG TCACGGGAGC GGGATGGCAA TGAGCGGGAC
351 AGCAGTGAGC CCGGGGATGA GGGTGGAGGG CGCAAGAGGC CTGTCCCTGA
401 TCAGACCTTG CAGCGCCGGG CAGGGTCAGG GACAGGGGTT GGGGCCATGC
451 TTGCTCGGGG CTCTGCTTCG CCCACAAAT CCTCTCCGCA GCCCTTGGTG
501 GCCACACCCA GCCAGCATCA CCAGCAGCAG CAGCAGCAGA TCAAACGGTC
551 AGCCCGCATG TGTGGTGAGT GTGAGGCATG TCGGCGCACT GAGGACTGTG
601 GTCACTGTGA TTTCTGTCTG GACATGAAGA AGTTCGGGGG CCCCAACAAG
651 ATCCGGCAGA AGTGCCGGCT GCGCCAGTGC CAGCTGCGGG CCCGGGAATC
701 GTACAAGTAC TTCCCTTCCT CGCTCTCACC AGTGACGCCC TCAGAGTCCC
751 TGCCAAGGCC CCGCCGGCCA CTGCCACCC AACAGCAGCC ACAGCCATCA
801 CAGAAGTTAG GCGCATCCG TGAAGATGAG GGGGCAGTGG CGTCATCAAC
851 AGTCAAGGAG CCTCCTGAGC CTACAGCCAC ACCTGAGCCA CTCTCAGATG
901 AGGACTTACC TCTGGATCCT GACCTGTATC AGGACTTCTG TGCAGGGGCC
951 TTTGATGACC ATGGCCCTGCC CTGGATGAGC GACACAGAAG AGTCCCCATT
1001 CCTGGACCCC GCGCTGCGGA AGAGGGCAGT GAAAGTGAAG CATGTGAAGC
1051 GTCGGGAGAA GAAGTCTGAG AAGAAGAAGG AGGAGCGATA CAAGCGGCAT
1101 CGGCAGAAGC AGAAGCACAA GGATAAATGG AAACACCCAG AGAGGGCTGA
1151 TGCCAAGGAC CTTGCGTCAC TGCCCCAGTG CCTGGGGCCC GGCTGTGTGC
1201 GCGCCGCCCA GCGCAGCTCC AAGTATTGCT CAGATGACTG TGGCATGAAG
1251 CTGGCAGCCA ACCGCATCTA CGAGATCCTC CCCAGCGCA TCCAGCAGTG
1301 GCAGCAGAGC CCTTGCAATG CTGAAGAGCA CGGCAAGAAG CTGCTCGAAC
1351 GCATTGCGCG AGAGCAGCAG AGTGCCCGCA CCCGCTTCA GGAATGGAA
1401 CGCCGATTCC ATGAGCTTGA GGCCATCATT CTACGTGCCA AGCAGCAGGC
1451 TGTGCCGAG GATGAGGAGA GCAACGAGGG TGACAGTGAT GACACAGACC
1501 TGCAATCTT CTGTGTTTCC TGTGGGCACC CCATCAACCC ACGTGTGACC
1551 TTGCGCCACA TGGAGCGCTG CTACGCCAAG TATGAGAGCC AGACGTCCTT
1601 TGGGTCCATG TACCCACAC GCATTGAAGG GGCCACACGA CTCTTCTGTG
1651 ATGTGTATAA TCCTCAGAGC AAAACATACT GTAAGCGGCT CCAGGTGCTG
1701 TGCCCGGAGC ACTCACGGGA CCCCAAAGT CCAGCTGACG AGGTATGCGG
1751 GTGCCCCCTT GTACGTGATG TCTTTGAGCT CACGGGTGAC TTCTGCCGCC
1801 TGCCCAAGCG CCAGTGCAAT CGCCATTACT GCTGGGAGAA GCTGCGCGCT
1851 GCGGAAGTGG ACTTGGAGCG CGTGCGTGTG TGGTACAAGC TGGACGAGCT
1901 GTTTGAGCAG GAGCGCAATG TGGCACAGC CATGACAAAC CGCGCGGGAT
1951 TGCTGGCCCT GATGCTGCAC CAGACGATCC AGCAGCATCC CCTCACTACC
2001 GACCTGCGCT CCAGTGCCGA CCGCTGAGCC TCCTGGCCCG GACCCTTAC
2051 ACCCTGCATT CCAGATGGGG GAGCCGCCCG GTGCCGTGTG GTCCGTTCCT
2101 CCACTCATCT GTTCTCCGG TTCTCCCTGT GCCCATCCAC CGGTGACCG
2151 CCCATCTGCC TTTATCAGAG GGAAGTGTCC CGTCGACATG TTCAGTGCCT
2201 GGTGGGGCTG CGGAGTCCAC TCATCCTTGC CTCTCTCCC TGGGTTTTGT
2251 TAATAAAATT TTGAAGAAAC CAAAAAATAA AAAAAAATAA AAAAAAATAA
2301 AAAAAAATAA AAAAAAATAA
```

BLAST Results

 Entry HS557771 from database EMBLEST:
 Human chromosome 18 clone 2 mRNA sequence.
 Score = 7582, P = 0.0e+00, identities = 1560/1598

Entry HS278337 from database EMBLEST:
 H.sapiens mRNA, expressed sequence tag ICRFp507H02194 (5')
 Score = 6339, P = 9.0e-281, identities = 1307/1347

Entry HS095149 from database EMBL:
 human STS WI-6941.
 Score = 1210, P = 2.2e-49, identities = 246/251

Medline entries

98449942:
 Identification and characterization of a family of mammalian methyl-CpG binding proteins.

9824997:
 Gene silencing by methyl-CpG-binding proteins.

Peptide information for frame 3

ORF from 57 bp to 2024 bp; peptide length: 656
 Category: similarity to known protein

```

1 MEGDGSDEPE PDAGEDSKSE NGENAPIYCI CRKPDINCFM IGCDCNNEWF
51 HGDCIRITEK MAKAIREWYC RECREKDPKL EIRYRHKKSR ERDGNERDSS
101 EPRDEGGGRK RPVPDPDLQR RAGSGTGVGA MLARGSASPH KSSPQPLVAT
151 PSQHHQQQQQ QIKRSARMCG ECEACRRTED CGHCDFCRDM KKEGPNKIR
201 QKCRLRQCQL RARESYKYFP SLSPTVTPSE SLPRPRRPLP TQQQPQPSQK
251 LGRIREDEGA VASSTVKEPP EATATPEPLS DEDLPLDPDL YQDFCAGAFD
301 DHGLFWMSDT EESPFLDPAL RKRAVKVKHV KRREKKSEKK KEERYKRHRQ
351 KQKHDKDKWH PERADAKDPA SLPQCLGPC VRPAQPSSKY CSDDCGMKLA
401 ANRIYEILPQ RIQQWQQSPC IAEHGGKLL ERIRREQQA RTRLQEMERR
451 FHELEAIIIR AKQQAVREDE ESNEGSDDT DLQIFCVSCG HPINPRVALR
501 HMERCYAKYE SOTSFGSMYP TRIEGATRLF CDVYNPQSKT YCKRLQVLCF
551 EHSRDPKVPA DEVCGCPLVR DVFELTGDFC RLPKRQCNRH YCWEKLRRAE
601 VDLEVRVRVWY KLDELFEQER NVRTAMTNRA GLLALMLHQT IQHDLPTTDL
651 RSSADR

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_4f17, frame 3

TREMBL:CEF52B11_4 gene: "F52B11.1"; Caenorhabditis elegans cosmid F52B11, N = 2, Score = 316, P = 8.8e-27

TREMBL:HSAB2331_1 gene: "KIAA0333"; Human mRNA for KIAA0333 gene, partial cds., N = 2, Score = 163, P = 2.8e-13

TREMBL:SPCC594_5 gene: "SPCC594.05c"; product: "putative transcriptional regulatory protein, phd finger containing"; S.pombe chromosome III cosmid c594., N = 3, Score = 168, P = 3.6e-12

TREMBL:AF072240_1 gene: "Mbd1"; product: "methyl-CpG binding protein MBD1"; Mus musculus methyl-CpG binding protein MBD1 (Mbd1) mRNA, complete cds., N = 2, Score = 189, P = 7.6e-11

>TREMBL:CEF52B11_4 gene: "F52B11.1"; Caenorhabditis elegans cosmid F52B11
 Length = 523

HSPs:

Score = 316 (47.4 bits), Expect = 8.8e-27, Sum P(2) = 8.8e-27
 Identities = 100/336 (29%), Positives = 167/336 (49%)

Score = 53 (8.0 bits), Expect = 8.8e-27, Sum P(2) = 8.8e-27
Identities = 24/100 (24%), Positives = 41/100 (41%)

Score = 48 (7.2 bits), Expect = 2.9e-26, Sum P(2) = 2.9e-26
Identities = 13/39 (33%), Positives = 19/39 (48%)

Pedant information for DKFZphtes3 4f17, frame 3

Report for DKFZphtes3 4f17.3

```

[FUNCAT]      99 unclassified proteins      [S. cerevisiae, YPL138c] 3e-10
[FUNCAT]      04.05.01.04 transcriptional control [S. cerevisiae, YNL097c] 2e-04
[PROSITE]     MYRISTYL      6
[PROSITE]     AMIDATION      2
[PROSITE]     CK2_PHOSPHO_SITE      8
[PROSITE]     TYR_PHOSPHO_SITE      3
[PROSITE]     GLYCOSAMINOGLYCAN      1
[PROSITE]     PKC_PHOSPHO_SITE      9
[KW]          All_Alpha
[KW]          LOW_COMPLEXITY      18.75 %
[KW]          COILED_COIL      4.57 %

```

```

SEQ      RAGSGTGVGAMLARGSASPHKSSPOPLVATPSQHHQQQQQIKRSARMGCEACEACRTED
SEG      .....xxxxxxxxx.....
PRD      cccccccceeeccccccccccccccccchhhhhhhhhhhhhhhhhccccccccccc
COILS

```

Prosites for DKFZphtes3 4f17.3

(No Pfam data available for DKFZphtes3 4f17.3)

DKF2phtes3_4f5

group: signal transduction

DKF2phtes3_4f5.3 encodes a novel 790 amino acid protein similar to beta-transducins.

The protein contains 3 WD-40 repeats, which are typical for the beta-transducin subunit of G-proteins. The beta subunits seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition. In addition, a Cytochrome C family heme-binding site signature is present. The protein is larger (790 amino acids) than the usual eukaryotic G-beta transducins (about 340 amino acids).

The new protein can find application in modulating/blocking G-protein-dependent pathways.

similarity to *S.pombe* "beta-transducin"

complete cDNA, EST hits

complete cds,

on genomic level encoded by HS313D11, at least 7 exons these exons

match

only partially with the predicted transcripts in HS313D11

Sequenced by AGOWA

Locus: /map="16p13.3"

Insert length: 3166 bp

No poly A stretch found, no polyadenylation signal found

```
1  GCGGCTTCC  GCGCGGCGG  TTCCGGACAA  CCGTGCGCTT  TTAGTAAAG
51  ATTGGGGTTC  GCGCGGGGGA  GAAGGGGTGC  CCCGGGCCCT  CTGGTTCTCG
101  TCCCGCAGCG  TCCGCTCCCC  CGCGCCACTG  CGCGCTCCC  AGGAACCCCTG
151  TACTCCGGGG  TCGCCGGCTT  CTCTCCTGCC  TCCGGTCCCG  CCAGACACCT
201  CGAGCTCCTT  AAGTAGCTCG  GTCCTTGACG  TCCCTCTGGG  CCCTTCCCGC
251  GTCTATCGCC  TGAGTCCCG  GGCCCTCTA  GCCCTCTGTT  CCCTCCCTC
301  TTTTGTTCCT  CCCTAGAGCC  CCGCCGCCCT  CAGGGCTGAC  AGTGTGGACG
351  GCGGGAGTCT  CCTCGCTCCC  CTGCTGGGAT  TGA CTGACCG  AGCGTTTAGT
401  GACTGCCCAG  ATCTGGCTGA  TGGGGGTACC  GAGAGGTGGC  CTGGGCCGGG
451  AATGTCCAGC  TAGAGTCTTC  CGTGGAAATC  AGACATGAAA  CTGACAGGCC
501  TAAGGGAAGC  TAGGAAGTCC  CCTCACCGCT  CAGCCAGGGT  GATGGGCTGG
551  ACTGACAGAC  TCCAGTGAAT  TTGAGCTTGC  CTGTGAGGCT  GATTGGCTGA
601  TAGACAGCCC  TGGATTGGCT  CACTAAGACT  GACCAGCCCG  GGACCAAGCA
651  GTTCTGGGGT  CCCAACCTGG  GTGGAAGGTC  TGAAGTGATG  ACCCACCAG
701  GCTGACCAGG  CCAGCCACCC  TCACTGACCT  CCTGACCCCT  GACCTCATCA
751  CCTGTGCAGC  CATGGAGAAG  ATGTCCCCTG  TGACCACAGC  CCTGGGTGGC
801  AGCGTGCTGA  CAGGCCGCAC  CATGCACTGC  CACCTGGATG  CTCCCGCCAA
851  TGCCATCAGT  GTGTGCCCGC  ACGCAGCCCA  GGTGGTCTGT  GCAGGCCGTA
901  GCATCTTCAA  GATCTATGCC  ATCGAGGAGG  AACAGTTCTG  GGAAAAGCTG
951  AACCTGCGTG  TGGGGCGCAA  GCCTTCGCTT  AACCTGAGCT  GTGCTGACGT
1001  GGTCTGGCAG  CAGATGGATG  AGAACCTGCT  GGCCACAGCA  GCCACCAATG
1051  GCGTGGTGGT  CACGTGGAAC  CTGGGCGCGC  CATCCCGCAA  CAAGCAGGAC
1101  CAGCTGTTC  CAGAACACAA  CGGCACGGTA  AACAAAGTCT  GCTTCCACCC
1151  CACCGAAGCC  CACGTGCTGC  TCAGTGGCTC  CCAGGATGGC  TTCATGAAGT
1201  GCTTTGACCT  CCGCAGAAAG  GACTCTGTCA  GCACCTTCTC  GGGCCAGTGG
1251  GAGAGCGTGC  GGGACGTGCA  GTTCAGTATC  CGGGACTACT  TCACCTTCGC
1301  CTCCACCTTT  GAGAACGGCA  ATGTGCAGCT  CTGGGACATC  CGGCGTCCCG
1351  ACCGGTGCGA  GAGGATGTTT  ACAGCCCA  ACGGACCCGT  CTTCTGCTGC
1401  GACTGGCACC  CCGAGGACAG  GGGCTGGTTG  GCCACTGGAG  GGCAGGACAA
1451  GATGGTGAAG  GTCTGGGACA  TGACCACGCA  CCGTGCCAAG  GAGATGCACT
1501  GTGTGCAGAC  CATCGCCTCG  GTGGCCCGTG  TGAAGTGGCG  GCCAGAGTGC
1551  CGCCACCACC  TGGCCACGTG  CTCCATGATG  GTGGACCACA  ACATCTATGT
1601  TTGGGACGTG  CGCCGGCCCT  TCGTGCCAGC  TGCCATGTTT  GAGGAACACC
1651  GAGACGTAC  CACGGGAATT  GCCTGGCGCC  ACCCCACAGA  CCCCTCCTTC
1701  CTGCTGTCTG  GCTCCAAGGA  CAGCTCGCTG  TGCCAGCACC  TGTTCCGCGA
1751  CGCCAGCCAG  CCCGTCGAGC  GCGCCAACCC  TGAGGGCCTC  TGCTACGGCC
1801  TCTTCGGGGA  CCTGGCCTTC  GCCGCCAAGG  AGAGCCTCGT  GGCTGCCGAG
1851  TCGGGGCGCA  AGCCCTACAC  TGGCGACCGG  CGCCACCCCA  TCTTCTTTAA
1901  GCGCAAGCTG  GACCTGCGCG  AGCCCTTCGC  AGGCCTCGCC  TCCAGTGCCC
1951  TCAGTGTCTT  TGAGACGGAG  CCAGTGGCG  GCGGCATGCG  CTGGTTTGTG
2001  GACACAGCTG  AGCGTTATGC  GCTGGCTGGC  CGGCCACTGG  CCGAGCTCTG
2051  TGACCAACAAC  GCAAAGGTGG  CTCGAGAGCT  TGGCCGCAAC  CAGGTGGCGC
2101  AAACGTGGAC  CATGCTGCGG  ATCATCTACT  GCAGCCCTGG  CCTAGTGCCC
2151  ACTGCAAACC  TCAACCACAG  TGTGGGCAAG  GGTGGCTCCT  GTGGCCTCCC
2201  GCTCATGAAC  AGTTTCAACC  TGAAGGATAT  GGCCCCAGGG  TTGGGCAGTG
2251  AGACGCGGCT  GGACCGCAGC  AAAGGAGATG  CACGGAGCGA  CACAGTTCTG
2301  CTCGACTCCT  CGGCCACACT  CATACCAAT  GAGGATAACG  AGGAAACCGA
2351  GGGCAGCGAC  GTACCTGCCG  ACTACCTGCT  GGGTGACGTG  GAAGGTGAGG
```

```

2401 AGGACGAGCT GTACCTGCTG GATCCGGAAC ACGCGCACCC CGAGGACCCT
2451 GAGTGGCTGC TGCCGCAGGA GGCCTTTCCG CTGCGCCACG AGATCGTGGA
2501 CACGCCTCCC GGACCCGAGC ACCTGCAGGA CAAGGCCGAC TCCCCGCACG
2551 TGAGCGGCAG CGAGGCGGAT GTGGCCTCCC TGGCCCCCGT GGA CTCTCC
2601 TTCTCGCTCC TGTCTGTCTC ACACGCGCTC TACGACAGCC GCCTGCCGCG
2651 CGACTTCTTC GGCCTGTCTG TGCGCGACAT GCTGCACTTC TACGTGAGC
2701 AGGGCGACGT GCAGATGGCT GTGTCTGTGC TCATCGTCCT GGGTGAACGG
2751 GTGCGCAAGG ACATCGACGA GCAGACCCAG GAGCACTGGT ACACCTCCTA
2801 CATCGACCTG CTGCAGCGCT TCCGCCTCTG GAACGTGTCC AACGAGGTGG
2851 TCAAGCTGAG CACCAGCCGC GCCGTCAGCT GCCTCAACCA GGCCTCCACC
2901 ACCCTGCACG TCAACTGCAG CCACTGCAAG CGGCCCATGA GCAGCCGGGG
2951 CTGGGTCTGC GACAGGTGCC ACCGCTGCGC CAGCATGTGT GCCGTCTGCC
3001 ACCAGGTAGT CAAGGGTCTC TTCGTGTGGT GCCAGGGCTG CAGCCACGGC
3051 GGCCACCTGC AGCACATCAT GAAGTGGCTG GAAGGCAGCT CCCACTGTCC
3101 CGCAGGCTGC GGCCACCTCT GCGAGTACTC CTGACGGGGC ATCTGCTGGG
3151 CTTGCCCGGG GGGCCG

```

BLAST Results

Entry HS313D11 from database EMBL:
 Human DNA sequence from cosmid 313D11 from a contig on the short arm of
 chromosome 16. Contains ESTs, STS and CpG islands.
 Score = 6238, P = 0.0e+00, identities = 1318/1391

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 762 bp to 3131 bp; peptide length: 790
 Category: similarity to known protein

```

1 MEKMSRVTTA LGGSVLTGRT MHCHLDAPAN AISVCRDAAQ VVVAGRSIFK
51 IYAIEEEQFV EKLNLRVGRK PSLNLSCADV VWHQMDENLL ATAATNGVVV
101 TWNLGRPSRN KQDQLFTEHK RTVNKVCFHP TEAHVLLSGS QDGFMKCFDL
151 RRKDSVSTFS GQSESVRDVQ FSIRDYFTFA STFENGTVQL WDIRRPDRCE
201 RMFTAHNGEV FCCDWHPEDR GWLATGGRDK MVKVVWDMTH RAKEMHCVQT
251 IASVARVKWR PECRHHLATC SMMVDHNIYV WDVRRPFVPA AMFEEHRDVT
301 TGIARHPPHD PSFLLSGSKD SSLCQHLFRD ASQPVERANP EGLCYGLFGD
351 LAFAAKESLV AAESGRKPYT GDRRHPIFFK RKLDPAPFFA GLASSALSVE
401 ETEPGGGGMR WFDVTAERYA LAGRPLAELC DHNAKVAREL GRNQVAQTWT
451 MLRIIYCSFG LVPTANLNHS VGKGGSCGLP LMNSFNKDM APGLGSETRL
501 DRSKGDARS DTVLLDSSATL ITNEDNEETE GSDVPADYLL GDVEGEDEEL
551 YLLDPEHAHP EDEPCVLPQE AFPLRHEIVD TPPGPEHLQD KADSPHVSGS
601 EADVASLAPV DSSFSLLSVS HALYDSRLPP DFFGVLRDM LHFYAEQGDV
651 QMAVSVLIVL GERVRKDIDE QTQEHWYTSY IDLLQRFRLW NVSNEVVKLS
701 TSRAVSCLNQ ASTTLHVNCS HCKRPMSSRG WVCDCRCHCA SMCVACHHVV
751 KGLFVWCQGC SHGGHLQHIM KWLEGSSHCP AGCGHLCEYS

```

BLASTP hits

Entry YDSB_SCHPO from database SWISSPROT:
 HYPOTHETICAL 93.2 KD TRP-ASP REPEATS CONTAINING PROTEIN C4F8.11 IN
 CHROMOSOME I. >TREMBL:SPAC4F8.11 gene: "SPAC4F8.11"; product:
 "beta-transducin"; S.pombe chromosome I cosmid c4F8.
 Score = 404, P = 3.0e-42, identities = 169/639, positives = 278/639

Entry PEX7 HUMAN from database SWISSPROT:
 PEROXISOMAL TARGETING SIGNAL 2 RECEPTOR (PTS2 RECEPTOR) (PEROXIN-7).
 >TREMBL:HSU76560.1 gene: "Pex7"; product: "peroxisome targeting signal
 2 receptor"; Human peroxisome targeting signal 2 receptor (Pex7) mRNA,
 complete cds. >TREMBL:HSU88871.1 gene: "HsPEX7"; product: "HsPex7p";
 Human HsPex7p (HsPEX7) mRNA, complete cds.
 Score = 220, P = 1.1e-15, identities = 62/244, positives = 107/244

Entry PEX7 MOUSE from database SWISSPROT:
 PEROXISOMAL TARGETING SIGNAL 2 RECEPTOR (PTS2 RECEPTOR) (PEROXIN-7).
 >TREMBL:MMU69171.1 product: "peroxisomal PTS2 receptor"; Mus musculus
 peroxisomal PTS2 receptor mRNA, complete cds.
 Score = 214, P = 5.3e-15, identities = 60/240, positives = 106/240

Entry ATAC2294.7 from database TREMBL:
 gene: "F11P17.7"; Arabidopsis thaliana chromosome I BAC F11P17 genomic
 sequence, complete sequence.
 Score = 232, P = 3.4e-14, identities = 68/260, positives = 120/260

Entry S66835 from database PIR:
 probable membrane protein YOL138c - yeast (*Saccharomyces cerevisiae*)
 >TREMBL:SCYOL138C_1 *S.cerevisiae* chromosome XV reading frame ORF
 YOL138c
 Score = 136, P = 2.5e-13, identities = 24/77, positives = 44/77

Alert BLASTP hits for DKFZphtes3_4f5, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_4f5, frame 3

 Report for DKFZphtes3_4f5.3

[LENGTH] 790
 [MW] 88207.10
 [pI] 6.05
 [HOMOL] SWISSPROT:YDSB_SCHPO HYPOTHETICAL 93.2 KD TRP-ASP REPEATS CONTAINING PROTEIN
 C4F8.11 IN CHROMOSOME I. 9e-44
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YOL138c] 5e-16
 [FUNCAT] 10.04.09 regulation of g-protein activity [S. cerevisiae, YBR195c] 3e-11
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YBR195c] 3e-11
 [FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YBR195c] 3e-11
 [FUNCAT] 09.13 biogenesis of chromosome structure [S. cerevisiae, YBR195c] 3e-11
 [FUNCAT] 04.05.01.07 chromatin modification [S. cerevisiae, YBR195c] 3e-11
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YCR072c beta-transducin family]
 3e-10
 [FUNCAT] 04.05.01.01 general transcription activities [S. cerevisiae, YBR198c
 TAF90 - TFIID subunit] 9e-09
 [FUNCAT] 04.01.04 rna processing [S. cerevisiae, YLL011w] 1e-07
 [FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae,
 YDL195w] 2e-07
 [FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL195w]
 2e-07
 [FUNCAT] 30.19 peroxisomal organization [S. cerevisiae, YDR142c] 4e-07
 [FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YDR142c]
 4e-07
 [FUNCAT] 08.10 peroxisomal transport [S. cerevisiae, YDR142c] 4e-07
 [FUNCAT] 08.01 nuclear transport [S. cerevisiae, YER107c] 4e-07
 [FUNCAT] 04.07 rna transport [S. cerevisiae, YER107c] 4e-07
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YER107c] 4e-07
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YGL003c] 5e-07
 [FUNCAT] 06.13 proteolysis [S. cerevisiae, YGL003c] 5e-07
 [FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YCR084c] 8e-07
 [FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YPR178w] 1e-06
 [FUNCAT] 03.13 meiosis [S. cerevisiae, YLR129w] 3e-06
 [FUNCAT] 03.25 cytokinesis [S. cerevisiae, YCR057c] 1e-05
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YCR057c]
 1e-05
 [FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation,
 palmitoylation, farnesylation and processing) [S. cerevisiae, YEL056w] 2e-04
 [FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YOR272w] 6e-04
 [SCOP] dlgoth_2.46.3.1.1 betal-subunit of the signal-transducing 5e-06
 [PIRKW] duplication 7e-10
 [PIRKW] signal transduction 7e-08
 [PIRKW] peroxisome 9e-06
 [PIRKW] heterotrimer 7e-08
 [PIRKW] GTP binding 7e-08
 [PIRKW] peroxisome biogenesis 9e-06
 [PIRKW] transmembrane protein 1e-14
 [SUPFAM] MS11 protein 7e-10
 [SUPFAM] WD repeat homology 1e-14
 [SUPFAM] GTP-binding regulatory protein beta chain 7e-08
 [SUPFAM] PRL1 protein 3e-08
 [SUPFAM] coatamer complex beta' chain 1e-06
 [PROSITE] CYTOCHROME_C 1
 [PROSITE] WD_REPEATS 3
 [PROSITE] MYRISTYL 10
 [PROSITE] AMIDATION 2
 [PROSITE] CAMP_PHOSPHO_SITE 2
 [PROSITE] CK2_PHOSPHO_SITE 11

```

[PROSITE]      TYR_PHOSPHO_SITE      1
[PROSITE]      PKC_PHOSPHO_SITE      7
[PROSITE]      ASN_GLYCOSYLATION     4
[PFAM]         WD_domain, G-beta repeats
[KW]           All_Beta
[KW]           3D
[KW]           LOW_COMPLEXITY        2.28 %

```

```

SEQ      MEKMSRVTTALGGSVLTRGTMHCHLDAPANAISVCRDAAQVVVAGRSIFKIYAIEEEQFV
SEG      .....
lgotB    .....

SEQ      EKLNLRVGRKPSLNLSCADVWHQMDENLLATAATNGVVVTWNLGRPSRNKQDQLFTEHK
SEG      .....
lgotB    .....TTCEEEEEETTTEEEET-TTTCEEE--EECCCC

SEQ      RTVNKVCFHPTAEHVLLSGSQDGMKCFDLRRKDSVSTFSGQSESVRDVQFSIRDYFTFA
SEG      .....
lgotB    CCEEEEEETT-TCEEEEEETTTEEEETTTEEEETTCCEEEEEETTTEEEEE

SEQ      STFENGVLQWDIIRPDRCEMFTAHNQPVFCCDWHPEDRGWLATGGRDKMVKVWDMTTH
SEG      .....
lgotB    E-ETTTTEEEETTTEEEEE-EECCCCCEEEEE-TTTCCEEEETTTEEEEC....

SEQ      RAKEMHCVQTIASVARVKWRPECRHHLATCSMMVDHNIYVWVRRPFVPAAMFEEHRDVT
SEG      .....
lgotB    .....

SEQ      TGIARHHPDPSFLLSGSKDSSLCQHLFRDASQPVERANPEGLCYGLFGDLAFAAKESLV
SEG      .....
lgotB    .....

SEQ      AAESGRKPYTGDRRHPIFFKRKLDPAEPFAGLASSALSVFETEPGGGMRWFVDTAERYA
SEG      .....
lgotB    .....

SEQ      LAGRPLAELCDHNAKVARELGRNQVAQTWMLRIIYCSPLVPTANLNHNSVGKGGSCGLP
SEG      .....
lgotB    .....

SEQ      LMNSFNLKDMAPGLGSETRLDRSKGDARSDTVLLDSSATLITNEDNEETEGSDVPADYLL
SEG      .....xxxx
lgotB    .....

SEQ      GDVEGEDELYLLDPEHAHPEDPECVLPQEAFLRHEIVDTPPGPEHLQDKADSPHVSQS
SEG      xxxxxxxxxxxxxxxx.....
lgotB    .....

SEQ      EADVASLAPVDSSFSLLSVSHALYDSRLPPDFFGVLVRDMLHFYAEQGDVQMAVSVLIVL
SEG      .....
lgotB    .....

SEQ      GERVRKDIDEQTQEHWYTSYIDLLQRFRLWNVSNEVVKLSTSRVSCNLQASTTLHVNC
SEG      .....
lgotB    .....

SEQ      HCKRPMSSRGWVCDRCHRCASMCVCHHVVKGLFVWCQGC SHGGLQHIMKWLEGSSHCP
SEG      .....
lgotB    .....

SEQ      AGCGHLCEYS
SEG      .....
lgotB    .....

```

Prosites for DKFZphtes3_4f5.3

PS00001	74->78	ASN_GLYCOSYLATION	PDOC00001
PS00001	468->472	ASN_GLYCOSYLATION	PDOC00001
PS00001	691->695	ASN_GLYCOSYLATION	PDOC00001
PS00001	718->722	ASN_GLYCOSYLATION	PDOC00001
PS00004	69->73	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	152->156	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	17->20	PKC_PHOSPHO_SITE	PDOC00005
PS00005	165->168	PKC_PHOSPHO_SITE	PDOC00005
PS00005	172->175	PKC_PHOSPHO_SITE	PDOC00005
PS00005	239->242	PKC_PHOSPHO_SITE	PDOC00005
PS00005	364->367	PKC_PHOSPHO_SITE	PDOC00005
PS00005	701->704	PKC_PHOSPHO_SITE	PDOC00005

PS00005	727->730	PKC_PHOSPHO_SITE	PDOC00005
PS00006	76->80	CK2_PHOSPHO_SITE	PDOC00006
PS00006	165->169	CK2_PHOSPHO_SITE	PDOC00006
PS00006	172->176	CK2_PHOSPHO_SITE	PDOC00006
PS00006	181->185	CK2_PHOSPHO_SITE	PDOC00006
PS00006	398->402	CK2_PHOSPHO_SITE	PDOC00006
PS00006	498->502	CK2_PHOSPHO_SITE	PDOC00006
PS00006	503->507	CK2_PHOSPHO_SITE	PDOC00006
PS00006	522->526	CK2_PHOSPHO_SITE	PDOC00006
PS00006	598->602	CK2_PHOSPHO_SITE	PDOC00006
PS00006	600->604	CK2_PHOSPHO_SITE	PDOC00006
PS00006	679->683	CK2_PHOSPHO_SITE	PDOC00006
PS00007	337->346	TYR_PHOSPHO_SITE	PDOC00007
PS00008	13->19	MYRISTYL	PDOC00008
PS00008	97->103	MYRISTYL	PDOC00008
PS00008	139->145	MYRISTYL	PDOC00008
PS00008	161->167	MYRISTYL	PDOC00008
PS00008	317->323	MYRISTYL	PDOC00008
PS00008	342->348	MYRISTYL	PDOC00008
PS00008	391->397	MYRISTYL	PDOC00008
PS00008	460->466	MYRISTYL	PDOC00008
PS00008	474->480	MYRISTYL	PDOC00008
PS00008	759->765	MYRISTYL	PDOC00008
PS00009	67->71	AMIDATION	PDOC00009
PS00009	364->368	AMIDATION	PDOC00009
PS00190	743->749	CYTOCHROME_C	PDOC00169
PS00678	90->105	WD_REPEATS	PDOC00574
PS00678	223->238	WD_REPEATS	PDOC00574
PS00678	269->284	WD_REPEATS	PDOC00574

Pfam for DKFZphtes3_4f5.3

HMM_NAME	WD domain, G-beta repeats		
HMM	*MrGHnnWVWCVaFSPDGrWFivSGSWDgTCRLWD*		
	++ HN++V C+ ++P+ R +++G++D+ +++WD		
Query	203	FTAHNGPVFCCDWHPEDRGWLATGGRDKMVKVWD	236

DKFZphtes3_4h6

group: intracellular transport/trafficking

DKFZphtes3_4h6 encodes a novel 622 amino acid protein with strong similarity to the kinesin light chain.

Kinesin is a microtubule-based motor protein that pulls vesicles or organelles towards the plus end of microtubules. Structural changes in the protein that drive motility are coupled to ATP binding and hydrolysis. The novel protein is similar to kinesin light chain, which is part of the functional kinesin holoenzyme tetrameric protein. The light chain has been proposed to function in coupling of cargo to the heavy chain or in the modulation of the ATPase activity of the heavy chain. The novel protein contains two kinesin light chain repeats and one RGD cell-attachment site.

The novel kinesin protein can find application in modulating the function of kinesin and modulating intracellular transport via/on microtubules.

strong similarity to Kinesin light chain

complete cDNA, complete cds, start at 150, EST hits (few)

Sequenced by AGOWA

Locus: unknown

Insert length: 2992 bp

Poly A stretch at pos. 2914, polyadenylation signal at pos. 2893

```

1  GCGGGGATGG AGGCGGCGGG ACCGGCTCGC GGGTGCGGGT CCGGGTGAAG
51  CGGGAGGCAG CCAGAGTCGG AGCCGGGCCC GAGCACCAGG CGCAGGCCCG
101 GCGCCCGCCT GCCCGCACCC TCGTCCTCAC AGACGCCACA GCCATGGCCA
151 TGATGGTGTT TCCGCGGGAG GAGAAGCTGA GCCAGGATGA GATCGTGCTG
201 GGCACCAAGG CTGTCATCCA GGGACTGGAG ACTCTGCGTG GGGAGCATCG
251 TGCCCTGCTG GCTCCTCTGG TTGCACCTGA GGCCGGCGAA GCCGAGCCTG
301 GCTCGCAGGA GCGCTGCATC CTCCTGCGTC GCTCCCTGGA AGCCATTGAG
351 CTTGGGCTGG GGGAGGCCCA GGTGATCTTG GCATTGTGCA GCCACCTGGG
401 GGCTGTAGAA TCAGAGAAGC AGAAGCTGCG GGCGCAGGTG CGGCGTCTGG
451 TGCAGGAGAA CCAGTGGCTG CGTGAGGAGC TGGCGGGGAC ACAGCAGAAG
501 CTGCAGCGCA GTGAGCAGGC CGTGGCCCCAG CTCGAGGAGG AGAAGCAGCA
551 CTTGCTGTTC ATGAGCCAGA TCCGCAAGTT GGATGAAGAC GCCTCCCTTA
601 ACGAGGAGAA GGGGGACGTC CCCAAGACA CACTGGATGA CCTGTCCCC
651 AATGAGGATG AGCAGAGCCC AGCCCTAGC CCAGGAGGAG GGGATGTGTC
701 TGGTCAGCAT GGGGGCTACG AGATCCCGGC CCGGCTCCGC ACCCTGCACA
751 ACCTGGTGAT CCAATACGCC TCACAGGGCC GCTACGAGGT AGCTGTGCCA
801 CTCTGCAAGC AGGCACTCGA AGACCTGGAG AAGACGTCA GGCACGACCA
851 CCCTGACGTT GCCACCATGC TGAACATCCT GGCACCTGGT TATCGGGATC
901 AGAACAAGTA CAAGGAGGCT GCCCACCTGC TCAATGATGC TCTGGCCATC
951 CGGGAGAAAA CACTGGGCAA GGACCACCCA GCGGTGGCTG CGACACTAAA
1001 CAACCTGGCA GTCCTGTATG GCAAGAGGGG CAAGTACAAG GAGGCTGAGC
1051 CATTGTGCAA GCGGGCACTG GAGATCCGGG AGAAGGTCTT GGGCAAGTTT
1101 CACCCAGATG TGGCCAAGCA GCTCAGCAAC CTGGCCCTGC TGTGCCAGAA
1151 CCAGGGCAAA GCTGAGGAGG TGGAAATATTA CTATCGGCGG GCACTGGAGA
1201 TCTATGCTAC ACGCCTCGGG CCCGATGACC CCAATGTGGC CAAGACCAAG
1251 AACAACTTGC CTTCTGTGTA CCTGAAGCAG GGCAAGTACC AGGATGCGGA
1301 GACCTGTGTA AAGGAGATCC TCACCCGCGC TCATGAGAAA GAGTTTGGCT
1351 CTGTCAATGG GGACAACAAG CCCATCTGGA TGCACGCAGA GGAGCGGGAG
1401 GAAAGCAAGG ATAAGCGCCG GGACAGCGCC CCTATGGGG AATACGGCAG
1451 CTGGTACAAG GCCTGTAAAG TAGACAGCCC CACAGTCAAC ACCACCTGCG
1501 GCAGCTTGGG GGCCCTATAC CGGCGCCAGG GCAAGCTGGA AGCCGCGCAC
1551 AACTAGAGG ACTGTGCCAG CCGTAACCGC AAGCAGGGTT TGGACCCCGC
1601 AAGCCAGACC AAGGTGGTAG AACTGCTGAA AGATGGCAGT GGCAGGCGGG
1651 GAGACCGCCG CAGCAGCCGA GACATGGCTG GGGGTGCCGG GCCTCGGTCT
1701 GAGTCTGACC TCGAGGACGT GGGACCTACA GCTGAGTGGG ATGGGGATGG
1751 CAGTGGCTCC TTGAGGCGCA GCGGTTCCTT TGGGAAACTC CGGGATGCCC
1801 TGAGGCGCAG CAGTGAGATG CTGGTAAAGA AGCTGCAGGG GGGCACCCCC
1851 CAGGAGCCCC CTAACCCAG GATGAAGCGG GCCAGTTCCT TCAACTTCCT
1901 CAACAAGAGC GTGGAAGAGC CGACCCAGCC TGGAGGCACA GGTCTCTCTG
1951 ACAGCCCGAC TCTCAGCTCC AGCTCCATGG ACCTCTCCCG ACGAAGCTCC
2001 CTGGTGGGCT AATGCTGAAG GGGCAGCCAG TCACCAGAGC GCCACCTGG
2051 CACACCCCCC TCACCCAGC CCTGCGCATG GGCTTGCTGC TTGTCCCGCC
2101 TGTCTCTCCC ACAGCCCTG TCTTTCTGT TCAATCTCAG GGTAACTTTC
2151 TCCCTTGTC TCTCAGCTG AGCCCTGGAG GCTGGGCTG CCCACTCCAG
2201 CTCATCCCTT TATTTATTC TTCCAGCAGG GCCCTCTTCC CTAGGTTCGG
2251 CCGAGCAGGA GGTGCCGGCT GGAGTCTCCA CCATAGACTC AGTGGCCTGG
2301 CCTCCCCAGA CCCAGAGGCC AAGAACACTA AGCACTCGCC GGCCCTTCGG
2351 CACCTCGCCG CTCCTCCCG ACTCAACCCG GCCGTTGCTT CTGTATATAG
2401 AGAAATAAGT TATTGGCCGC GCGCTCCCT TCAGTCCAGG GTACTACCCG

```

```

2451 GGCCTCCCTT CGTCCCTCTT CTAGTGGTAC CGCCAGGCC TTAATCACCC
2501 CCATTCCGTG CGGTGGTATC TCCCAGGCTC TACATTCTCG GGAGCGGCGC
2551 CTCCCAAGGG GGTCTTGGGA CCTTCTCGCG CTCCTCTGGG CCTCTGAGGG
2601 ATGCGTCCTA CCCGCGCCAT CGCCCGCTGG CCCAGGACGG GGACCTCCCC
2651 TTAGTCCGTC CTCCCACCGC CGGGCCCTGC CCCGCATCCC GGCCTTATGC
2701 ACTGCCCTCT CCACCCGGCC CGGCCAGGC ACGGCCGACC CCGCCCCGGG
2751 CACCGCCAC CGAGCCATCC TGCCTCGCCT CCCCCACGC CTGCAGCTTC
2801 TCGCGAGGGG CGGCGACGGT CCCCTGGTGG CAGGAGGGGC TCCCCTGT
2851 GCGGGTGAGG CGGCTGCTCT CTATTTTCAG ATGTTGCTGT AGAAATAAAG
2901 ACGGTTTAAA TCTGAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2951 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AA

```

BLAST Results

No BLAST result

Medline entries

98288268:

Two kinesin light chain genes in mice. Identification and characterization of the encoded proteins.

Peptide information for frame 3

ORF from 144 bp to 2009 bp; peptide length: 622
 Category: strong similarity to known protein
 Prosite motifs: RGD (502-505)
 KINESIN_LIGHT (223-265)
 KINESIN_LIGHT (265-307)

```

1 MAMMVFPREE KLSQDEIVLG TKAVIQGLET LRGEHRALLA PLVAPEAGEA
51 EPGSQERCIL LRRSLEAIEL GLGEAQVILA LSSHLGAVES EKQKLRAQVR
101 RLVOENQWLR EELAGTQOKL QRSEQAVAQL EEEKQHLLFM SQIRKLDEDA
151 SPNEEKGVDP KDTLDDLFPN EDEQSPAPSP GGGDVSGQH GYEIPARLRT
201 LHNLIQIYAS QGRYEVAVPL CKQALDLEK TSGHDHPDVA TMLNILALVY
251 RDQNKYKEAA HLLNDALAIR EKTGKDHFA VAATLNNLAV LYGKRKYKE
301 AEPLCKRALE IREKVLGKFH PDVAKQLSNL ALLCQNGKA EEVEYYRRA
351 LEIYATRLGP DDPNVAKTKN NLASCYLKQG KYQDAETLYK EILTRAHEKE
401 FGSVNGDNKP IWMHAEEREE SKDKRRDSAP YGEYGSWYKA CKVDSPTVNT
451 TLRSLGALYR RQKLEAAHT LEDCASNRK QGLDPASQTK VVELLKDGSG
501 RRGDRRSSRD MAGGAGPRSE SDLEDVGPTA EWNGDGSGL RRSFSFGKLR
551 DALRRSSEML VKKLQGGTPQ EPPNPRMKRA SSLNFLNKS SV EPTQPGGTG
601 LSDSRTLSSS SMDLSRRSSL VG

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_4h6, frame 3

TREMBL:AF055666_1 gene: "Klc2"; product: "kinesin light chain 2"; Mus musculus kinesin light chain 2 (Klc2) mRNA, complete cds., N = 1, Score = 2824, P = 4e-294

PIR:I53013 kinesin light chain - human, N = 1, Score = 1927, P = 4.5e-199

PIR:C41539 kinesin light chain C - rat, N = 1, Score = 1919, P = 3.2e-198

SWISSPROT:KNLC_RAT KINESIN LIGHT CHAIN (KLC)., N = 1, Score = 1919, P = 3.2e-198

>TREMBL:AF055666_1 gene: "Klc2"; product: "kinesin light chain 2"; Mus musculus kinesin light chain 2 (Klc2) mRNA, complete cds.
 Length = 599

HSPs:

Score = 2824 (423.7 bits), Expect = 4.0e-294, P = 4.0e-294
 Identities = 558/598 (93%), Positives = 572/598 (95%)

```

Query:   1 MAMMVFPREEKLSQDEIVLGTKAVIQGLETLRGEHRALLAPLVAPEAGEAEPGSQERCIL 60
          MA MV PREEKLSQDEIVLGTKAVIQGLETLRGEHRALLAPL + EAGEAEPGSQERC+L
Sbjct:   1 MATMVLPREEKLSQDEIVLGTKAVIQGLETLRGEHRALLAPLASHEAGEAEPGSQERCLL 60

Query:  61 LRRSLEAIELGLGEAQVILALSSHLGAVESEKQKLRAQVRRVLVQENQWLREELAGTQQKL 120
          LRRSLEAIELGLGEAQVILALSSHLGAVESEKQKLRAQVRRVLVQENQWLREELAGTQQKL
Sbjct:  61 LRRSLEAIELGLGEAQVILALSSHLGAVESEKQKLRAQVRRVLVQENQWLREELAGTQQKL 120

Query: 121 QRSEQAVAQLEEEKQHLLFMSQIRKLDE DASPNEEKGDVDPKDTLDDLPNEDEQSPAPSP 180
          QRSEQAVAQLEEEKQHLLFMSQIRKLDE P EEKGDVDPK+LDDLPNEDEQSPAPSP
Sbjct: 121 QRSEQAVAQLEEEKQHLLFMSQIRKLDE-MLPQEEKGDVDPKDSLDDLPNEDEQSPAPSP 179

Query: 181 GGGDVSGQHGGYEIPARLRTLHNLVIQYASQGRYEVAVPLCKQALEDLEKTSBGHDHPDVA 240
          GGGDV+ QHGGYEIPARLRTLHNLVIQYASQGRYEVAVPLCKQALEDLEKTSBGHDHPDVA
Sbjct: 180 GGGDVAAQHGGYEIPARLRTLHNLVIQYASQGRYEVAVPLCKQALEDLEKTSBGHDHPDVA 239

Query: 241 TMLNILALVYRDQNKYKEAAHLLNDALAIREKTLGKDHPAVAATLNNLAVLYGKRKYKE 300
          TMLNILALVYRDQNKYK+AAHLLNDALAIREKTLGKDHPAVAATLNNLAVLYGKRKYKE
Sbjct: 240 TMLNILALVYRDQNKYKDAHLLNDALAIREKTLGKDHPAVAATLNNLAVLYGKRKYKE 299

Query: 301 AEPLCKRALEIREKVLGKFHPDVAKQLSNLALLCQNGKAEVEYYYRRALEIYATRLGP 360
          AEPLCKRALEIREKVLGKFHPDVAKQLSNLALLCQNGKAEVEYYYRRALEIYATRLGP
Sbjct: 300 AEPLCKRALEIREKVLGKFHPDVAKQLSNLALLCQNGKAEVEYYYRRALEIYATRLGP 359

Query: 361 DDPNVAKTKNNLASCYLKQKGYQDAETLYKEILTRAHEKEFGSVNGDNKPIWMHAEEREE 420
          DDPNVAKTKNNLASCYLKQKGYQDAETLYKEILTRAHEKEFGSVNG+KPIWMHAEEREE
Sbjct: 360 DDPNVAKTKNNLASCYLKQKGYQDAETLYKEILTRAHEKEFGSVNGENKPIWMHAEEREE 419

Query: 421 SKDKRRDSAPYGEYSWKACKVDSPTVNTTLRSLGALYRRQGGKLEAAHTLEDCASRNK 480
          SKDKRRD P EYGSWKACKVDSPTVNTTLR+LGALYR +GKLEAAHTLEDCASR+RK
Sbjct: 420 SKDKRRDRPM-EYGSWKACKVDSPTVNTTLRLTALYRPEGKLEAAHTLEDCASRSRK 478

Query: 481 QGLDPASQTKVVELLKDGSRRGDRSSRD MAGGAGPRSESDLEDVGP TAEWNGDGSGL 540
          QGLDPASQTKVVELLKDGSRR G RR SRD+AG P+SESDLE+ GP AEW+GDGSGL
Sbjct: 479 QGLDPASQTKVVELLKDGSRR-GHRRGSRDVAG---PQSESDLEESGPAAEWSGDGSGL 534

Query: 541 RRSFSFGKLRDALRRSSEMLVKKLQGGTPQEPNPRMKRASSLNFLNKSVEEPTQPGG 598
          RRSFSFGKLRDALRRSSEMLV+KLQGG PQEP N RMKRASSLNFLNKSVEEP QPGG
Sbjct: 535 RRSFSFGKLRDALRRSSEMLVRKLQGGGPQEP-NSRMKRASSLNFLNKSVEEPVQPGG 591

```

Pedant information for DKF2phtes3_4h6, frame 3

Report for DKF2phtes3_4h6.3

```

[LENGTH]      622
[MW]           68934.82
[pI]           6.72
[HOMOL]       TREMBL:AF055666_1 gene: "Klc2"; product: "kinesin light chain 2"; Mus musculus
kinesin light chain 2 (Klc2) mRNA, complete cds. 0.0
[BLOCKS]      BL00927C Trehalase proteins
[BLOCKS]      BL01160I Kinesin light chain repeat proteins
[BLOCKS]      BL01160H Kinesin light chain repeat proteins
[BLOCKS]      BL01160G Kinesin light chain repeat proteins
[BLOCKS]      BL01160F Kinesin light chain repeat proteins
[BLOCKS]      BL01160E Kinesin light chain repeat proteins
[BLOCKS]      BL01160D Kinesin light chain repeat proteins
[BLOCKS]      BL01160C Kinesin light chain repeat proteins
[BLOCKS]      BL01160B Kinesin light chain repeat proteins
[BLOCKS]      BL01160A Kinesin light chain repeat proteins
[SUPFAM]      tetratricopeptide repeat homology 1e-07
[PROSITE]     RGD 1
[PROSITE]     MYRISTYL 8
[PROSITE]     KINESIN_LIGHT 2
[PROSITE]     AMIDATION 2
[PROSITE]     CAMP_PHOSPHO_SITE 5
[PROSITE]     CK2_PHOSPHO_SITE 11
[PROSITE]     TYR_PHOSPHO_SITE 3
[PROSITE]     PKC_PHOSPHO_SITE 7
[PROSITE]     ASN_GLYCOSYLATION 2
[PFAM]        Kinesin light chain repeat
[KW]          All Alpha
[KW]          LOW_COMPLEXITY 12.54 %
[KW]          COILED_COIL 4.98 %

```


Prosite for DKFZphtes3 4h6.3

888

PS00006	568->572	CK2_PHOSPHO_SITE	PDOC00006
PS00006	589->593	CK2_PHOSPHO_SITE	PDOC00006
PS00006	610->614	CK2_PHOSPHO_SITE	PDOC00006
PS00007	339->346	TYR_PHOSPHO_SITE	PDOC00007
PS00007	339->347	TYR_PHOSPHO_SITE	PDOC00007
PS00007	424->432	TYR_PHOSPHO_SITE	PDOC00007
PS00008	71->77	MYRISTYL	PDOC00008
PS00008	86->92	MYRISTYL	PDOC00008
PS00008	182->188	MYRISTYL	PDOC00008
PS00008	187->193	MYRISTYL	PDOC00008
PS00008	402->408	MYRISTYL	PDOC00008
PS00008	482->488	MYRISTYL	PDOC00008
PS00008	598->604	MYRISTYL	PDOC00008
PS00008	600->606	MYRISTYL	PDOC00008
PS00009	292->296	AMIDATION	PDOC00009
PS00009	499->503	AMIDATION	PDOC00009
PS00016	502->505	RGD	PDOC00016
PS01160	223->265	KINESIN_LIGHT	PDOC00893
PS01160	265->307	KINESIN_LIGHT	PDOC00893

Pfam for DKFZphtes3_4h6.3

HMM_NAME Kinesin light chain repeat

HMM *RALEDREKtLGHDHPDVatMLNNLALvCRNQNKYeEvenYYN*
+ALED+EKT+GHDHPDVATMLN+LALV+R+QNKY+E++ ++N

Query 223 QALEDLKETSghdHPDVatMLNlALVYRDQNKYKEAAHLN 264

50.46 265 306 1 42 dkfzphes3_4h6.3 strong similarity to Kinesin light chain
Alignment to HMM consensus:

Query *RALEDREKtLGHDHPDVatMLNNLALvCRNQNKYeEvenYYN*
AL +REKTLG DHP VA LNNLA+++ ++KY+E+E + +
dkfzphes3 265 DALAIREKTLGKDHPAVAATLNNLAVLYGKRGKYKEAEPLCK 306

Query 348 1 42 dkfzphes3_4h6.3 strong similarity to Kinesin light chain
Alignment to HMM consensus:

HMM *RALEDREKtLGHDHPDVatMLNNLALvCRNQNKYeEvenYYN*
RALE+REK+LG HPDVA++L+NLAL+C+NQ+K EEVE YY+

Query 307 RALEIREKVLGKFHPDVAKQLSNLALLCQNGKAEEVEYYR 348

39.10 349 390 1 42 dkfzphes3_4h6.3 strong similarity to Kinesin light chain
Alignment to HMM consensus:

Query *RALEDREKtLGHDHPDVatMLNNLALvCRNQNKYeEvenYYN*
RALE+ LG D P+VA+ NNLA + Q+KY+++E +Y+
dkfzphes3 349 RALEIYATRLGPDDPNVAKTKNNLASCYLKQGKYQDAETLYK 390

DKFZphtes3_4o19

group: testes derived

DKFZphtes3_4o19 encodes a novel 1180 amino acid protein with weak similarity to human megakaryocyte stimulating factor and human mucin.

The novel protein contains a cytochrome c family heme-binding site signature.
No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to megakaryocyte stimulating factor and mucin

complete cDNA, complete cds, EST hits (few)

Sequenced by AGOWA

Locus: unknown

Insert length: 3767 bp

Poly A stretch at pos. 3757, polyadenylation signal at pos. 3737

```
1 GGCTAGGTTT AGCTTCAGGG GCAGCCCAGG GCAGTGTGTC TGCATATTGC
51 ATGGATGAAA GGCTGAAGGC TGCTCCTCT TGCAGGCTGG CTTCTGAGAT
101 TGCACCTTCT TCTCCTGCTA CTCCTCCAAA TCTATGACCC TCAAGGCAG
151 AGCTGACCTG TCCGGTAATC AAGGCAATGC AGCCGGCCGC CTAGCTACAG
201 TTCACGAGCC AGTTGTCACC CAGTGGGCGG TGCATCCTCC AGCCCCCGCT
251 CACCCAGTCT TCCTGGACAA AATGGAGAAA GCGCCTCCAC AGCCCCAGCA
301 CGAGGGCGTC AAGTCCAAGG AGCATCTTCC GCAACAGCCT GCCGAAGGCA
351 AGACGGCGTC CCGCCGCGTC CCACGCCTCC GGGCTGTGGT CGAGAGCCAG
401 GCCTTCAAGA ACATCCTGGT AGACGAGATG GACATGATGC ACGCCCGTGC
451 AGCCACGCTC ATCCAAGCCA ACTGGAGGGG CTATTGGCTC CGGCAGAAGC
501 TGATTTCCTA GATGATGGCG GCCAAGGCCA TCCAGGAGGC CTGGCGGCGC
551 TTCACCAAGA GACACATCCT TCACTCCAGC AAGTCGTTGG TAAAGAAAAC
601 GAGGGCGGAG GAGGGGGACA TACCTTATCA CGCCCCACAG CAGGTGCGCT
651 TCCAGCATCC GGAAGAGAAC CGCCTTCTGT CCCC GCCCAT CATGGTGAAC
701 AAGGAGACCC AGTTCCCTTC CTGTGACAAT CTGGTCTCTC GCAGACCCCA
751 GTCTGCCCCC CTCTGCGAGC CCCCAGCAGC TCAGGGTACC CCAGAGCCCT
801 GTGTGCGAGG TCCTCATGCT GCCAGAGTCC GGGGGCTGGC CTTCTGCCA
851 CACCAGACGG TCACCATCAG ATTTCCCTGC CCAGTGAGTT TGGACGCAAA
901 ATGCCAGCCA TGCTGCTGA CCAGAACCAT CAGAAGCACC TGCTCGTCC
951 ACATAGAGGG TGACTCAGTG AAGACCAAAC GTGTAAGTGC CCGGACCAAC
1001 AAAGCCAGGG CTCCGGAGAC ACCATTGTCC AGAAGGTATG ACCAGGCAGT
1051 TACGAGACCA TCCAGAGCCC AAACCCAGGG CCCTGTGAAA GCAGAGACCC
1101 CCAAGGCCCC CTTCCAGATA TGTCCAGGGC CCATGATCAC CAAGACTCTA
1151 CTCCAGACAT ATCCAGTGGT CTCCTGACCC CTGCCACAGA CATATCCAGC
1201 GTCCACGATG ACCACCACCC CACCCAAGAC TAGCCAGTT CCCAAAGTAA
1251 CAATAATCAA GACCCAGGCC CAGATGTATC CGGGGCCAC AGTGACCAAA
1301 AACTGCACCT ACACATGCCC CATGCCACA ATGACCAAGA TCCAGGTACA
1351 CCCCACAGCC TCCAGAACTG GCACCCACG GCAGACATGC CCTGCGACCA
1401 TCACGGCAAA GAACCGACCT CAGGTTTCCC TTCTGGCTTC CATCATGAAG
1451 AGCTTGCCCC AGGTATGCCC GGGGCTGCG ATGGCAAGA CCCCACCCCA
1501 GATGCACCCG GTCACCACCC CAGCCAAAAA CCCATTGCAA ACATGTCTGT
1551 CAGCCACAAT GTCCAAGACT TCATCCCAGA GGAGCCAGT TGGGTGACC
1601 AAGCCCTCAC CCCAGACCCG CCTGCCAGCC ATGATAACCA AGACCCAGC
1651 CCAGTTAGCG TCGGTGGCCA CCATCCTCAA GACTCTGTGT CTGGCCTCTC
1701 CAACAGTGGC AAATGTCAAG GCTCCACCCC AAGTGGCGGT AGCAGCCGGA
1751 ACTCCCAACA CCTCAGGCTC CATCCATGAG AACCACCCA AGGCCAAGGC
1801 CACCGTGAAT GTGAAGCAGG CTGCAAAAGT GGTGAAAGCC TCATCCCCCT
1851 CCTATTTGGC TGAGGGGAAG ATCAGGTGCC TGGCTCAACC ACATCCGGGA
1901 ACTGGGGTCC CCAGGGCTGC AGCTGAGCTT CCTTTGGAAG CCGAGAAAAT
1951 CAAGACTGGC ACCCAGAAAC AGGCGAAAAC AGACATGGCA TTAAAGACCA
2001 GTGTGGCAGT GGAATGGCT GGGGCTCCAT CTTGGACAAA AGTTGCTGAG
2051 GAAGGGGACA AGCCACCTCA CGTGTATGTG CCTGTAGACA TGGCTGTAC
2101 CCTGCCCGG GGACAGCTGG CTGCCCCACT GACCAATGCC TCATCCCAGA
2151 GACATCCACC CTGCTGTCC CAGAGACCAC TGGCCGCCCC GCTGACCAAG
2201 GCCTCATCTC AGGGACATCT GCCCACTGAG CTGACCAAGA CCCCATCCCT
2251 GGCCCATCTG GACACCTGTC TGAGCAAGAT GCATTCCCAG ACACATCTGG
2301 CCACAGGTGC CGTGAAGGTC CAGTCCCAAG CGCCTCTAG CACCTGTCTG
2351 ACCAAGACGC AGTCCCGGGG GCAGCCGATC ACAGACATAA CCACGTGCCT
2401 CATCCCAGCG CACCAGGCTG CTGATCTCAG CAGCAACACC CACTCCCAGG
2451 TGCTCCTAAC AGGGTCCAAG GTGTCCAACC ACGCTGCCA GCGCCTCGGT
2501 GGCCTCAGCG CCCACCTCTG GGCCAAGCCA GAGGACAGAC AGACCCAGCC
2551 ACAGCCCCAC GGACACGTGC CGGGGAAGAC CACTCAGGGG GGACCATGCC
2601 CGGCAGCCTG TGAGGTCCAG GGTATGCTGG TGCCGCCGAT GGCACCCACC
```

```

2651 GGCCATTCCA CATGCAACGT TGAGTCCTGG GGAGACAACG GAGCCACACG
2701 TGCCAGCCCA TCAATGCCCG GCCAGGCGGT GCCCTGCCAG GAGGACACGG
2751 GCGCCGCGGA CGCTGGTGTG GTTGGTGGCC AATCGTGGAA CCGCGCATGG
2801 GAGCCAGCCA GGGGTGCTGC GTCCTGGGAC ACCTGGCGCA ACAAGGCGGT
2851 GGTGCCTCCC AGGCGGTCCG GGGAGCCAAT GGTGTCCATG CAGGCTGCAG
2901 AGGAGATCCG CATCTCGCA GTGATCACTA TCCAGGCGGG CGTCCGTGGC
2951 TACCTGGCGC GTCGCAGGAT CCGGCTGTGG CACCGGGGGG CCATGGTCAT
3001 CCAAGCTACT TGGCGCGGCT ACCGTGTGCG GCGGAACCTG GCACACCTCT
3051 GCAGAGCCAC CACGACCATC CAGTCTGCCT GGCGCGGCTA CAGCACCCGC
3101 CGGGACCAAG CCCGGCACTG GCAGATGCTC CACCCGCTCA CGTGGGTGGA
3151 GCTGGGCAGC CGGGCGGGG TCATGTCTGA CCGAAGCTGG TTCCAGGATG
3201 GCAGAGCCAG GACAGTATCT GACCATCGCT GCTTCCAGTC CTGCCAGGCA
3251 CACGCTTGCA GCGTCTGCCA CTCCCTGAGC TCCAGGATCG GGAGCCCGCC
3301 CAGCGTGGTG ATGCTAGTGG GCTCCAGCCC TCGCACCTGT CATACCTGTG
3351 GACGCACACA GCCCACCCTG GTGGTGCAGG GCATGGGCCA GGGCACTGAG
3401 GGCCCGGGG CAGTGTCTTG GGCCTCCGCC TACCAGCTGG CTGCCCTGAG
3451 TCCCAGGAG CCGCATCGCC AGGACAAAGC GGCCACAGCC ATCCAGTCCG
3501 CCTGAGGGG CTTTAAGATC CGCCAGCAGA TGAGGCAGCA GCAATGGCA
3551 CGAAGATAG TTCAAGCCAC CTGGCGAGGC CACCATACCC GGAGCTGTCT
3601 GAAGAACACA GAGGCGCTCT TGGGACCAGC AGACCCTCG GCCAGCTCAC
3651 GGCACATGCA TTGGCCTGGC ATCTAGGACC CTGGCTCCCT GCAGTGGGGA
3701 CTTCTGTTGA GGCACCTCAT GCTCTCTGGG TCTAATGAAT AAAGTCTCTC
3751 ACAGCCTAAA AAAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 134 bp to 3673 bp; peptide length: 1180
 Category: similarity to known protein

```

1 MTLQGRADLS GNQNAAGRL ATVHEPVVTQ WAVHPPAPAH PSLLDKMEKA
51 PPQPQHEGLK SKEHLPQOPA EGKTASRRVP RLRAVVESQA FKNILVDEMD
101 MMHARAATLI QANWRGYWLR QKLISQMAA KAIQEAWRRF NKRHILHSSK
151 SLVKKTAAEE GDIPYHAPQQ VRFQHPENR LLSPPIMVKN ETQFPSCDNL
201 VLCRPQSSPL LQPPAAQGTG EPCVQGPAA RVRGLAFLPH QTVTIRFPCP
251 VSLDAKQCPC LLTRTIRSTC LVHIEGDSVK TKRVSARTNK ARAPETPLSR
301 RYDQAVTRPS RAQTQGPVKA ETPKAPFQIC PGPMITKTL QTYPVVSVTL
351 PQTYPASTMT TTPPKTSVPV KVTIIKTPAQ MYPGPTVTK APHTCPMPTM
401 TKIQVHPTAS RTGTPRQTC ATITAKNRPO VSLLASIMKS LPQVCPGPAM
451 AKTPQMHPV TTPAKNPLQ CLSATMSKTS SQRSPVGVTK PSPQTRLPM
501 ITKTPAQLRS VATILKTLCL ASPTVANVKA PPQVAVAAGT PNTSGSIHEN
551 PPKAKATVNV KQAAKVVKAS SPSYLAEGKI RCLAQPHPGT GVPRAAAELP
601 LEAEKIKTGT QKQAKTDMAF KTSVAVEMAG APSWTKVAAE GDKPPHVVYP
651 VDMAVTLPRG QLAAPLTNAS SQRHPPCLSQ REPLAAPTKA SSQGHLPTEL
701 TKTPSLAHL TCLKMHSQT HLTGAVKVQ SQAPLATCLT KTQSRGPIT
751 DITCLIPAH QAADLSSNTH SQVLLTGSKV SNHACQRLGG LSAPPWAKPE
801 DRQTQPPHGS HVPKTTQGG PCPAACEVQG MLVPPMAPTG HSTCNVESWG
851 DNGATRAQPS MPGQAVPCQE DTGPADAGVV GGQSWNRAWE PARGAASWD
901 WRNKAVVPPR RSSEPVMVSMQ AAEEIRILAV ITIQAGVRGY LARRRIRLWH
951 RGAMVIQATW RGYRVRRNLA HLCRATTTIQ SAWRGYSTRR DQARHWQMLH
1001 PVTWVELGSR AGVMSDRSWF QDGRARTVSD HRCFQSCQAH ACSVCHSLSS
1051 RIGSPSPVVM LVGSSPRTCH TCGRTQPTRV VQGMGQGTG PGAVSWASAY
1101 QLAALSPPRP HRQDKAATAI QSAWRGFKIR QMRQQQMAA KIVQATWRGH
1151 HTRSLKNTKE ALLGPADPSA SSRHMHWPFI

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_4o19, frame 2

TREMBL:HSU70136_1 product: "megakaryocyte stimulating factor"; Human
 megakaryocyte stimulating factor mRNA, complete cds., N = 2, Score =
 242, P = 9.6e-16

Query:	281	TKRVSARTNKARAPETPLSRRYDQAVTRPSRAQTQGPVKAE--TPKAPFQICPGPMITKT	338
		TK+ + K AP TP + A T P + P K TP+ P P + T	
Sbjct:	597	TKKPAPTAPKEPAPTTTPK----ETAPTTPKKLPTTPEKLAPTTPEKPAPTTPEELAPTT	652
Query:	339	LLQTYPVVSVTLPPQTYPASTMTTTPPKTSPV-PKVITIKTPAQMPGPGTVTK-TAPHTCP	396
		+ P T P + TP + +P PK TP + P PT K TAP T P	
Sbjct:	653	PEEPTPTTPEEPAPTTPKAAAPNTKPEPAPTTKPEPAPTTPEK--PAPTTPKETAP-TTP	709
Query:	397	M---PTMTKIQVHPTASRTGTPRQTCPATITAKNRQVSLASIMKSLPQVCPGPMAMAKT	453
		PT K + PT + P++ PT + S + K P GA T	
Sbjct:	710	KGTAPTTLK-EPAPTTPKKPAPKELAPTT----TKEPTSTTSD--KPAPTTPKGTAPT-T	761
Query:	454	PPQMHPVTTPAKNPLQTCLSATMKTSSQRSRSPVGVTKPSQTRLPAMITKTPAQLRSVAT	513
		P + P TTP K P T T T + +P KP+P+ P TK P S	
Sbjct:	762	PKEPAP-TTP-KEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTT-TKGPTSTTSDKP	818
Query:	514	ILKTLCLASPTVANVKAPPQVAVAAGTPNTSGSIHENPPKAKATVNV----QAAKVVKA	569
		T +PT AP A PT E PP + V+ K+ + K+	
Sbjct:	819	APTTPKETAPTTKPEAPATTTPKKPA--PTTP---ETPPPTTSEVSTPTTTTKEPTTIHKS	872
Query:	570	---SSPSYLAEGKIRCLAQPHPGTGVPRAAAELEAEIKITGTQKQAKTDMAFKTSVAV	626
		S+P AE + L GVP + P + T T K T+ +T+	

Sbjct: 873 PDESTPELSAETPKALENSPKKEGVP--TTKTPAATKPEMTTAKDKTTERDLRTTPET 930

Query: 627 EMAGAPSWTK-VAEEGDKPPHVYPVDMAVTLPRGQLAAPLTNASSQRHPCLSQRLAA 685
A AP TK A +K + +T Q+ + T ++ L LA

Sbjct: 931 TTA-APKMTKETATTTTEKT-----TESKITATTTQVSTTTTQDTPFKITTLKTTTLAP 983

Query: 686 PLTKASSQGHLPTELTKTPSLAHLDTCLSKMHSQTHLATGAVKVQS-----QAPLATCLT 740
+T + + TE+ P +T K + AT K Q + P +T

Sbjct: 984 KVT-TTKKTITTTTEIMNKPE----ETAKPKDRATNSKAT-TPKPQKPTKAPKKPTSTKKP 1037

Query: 741 KTQSR-GQPITDIT---TCLIPAHQAADLSSNTHSQVLLTGSKVSNHACQRLGGLSAPP 795
KT R +P T T T +P + Q ++ N + S

Sbjct: 1038 KTMPRVRKPKTTTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDA 1097

Query: 796 W-AKPEDRQTQPPHGHVPGKTTQGGPCPAACEVQGMVPPMAPTGHSTCN 845
A+ E +PH +P T P OG+++ PM + CN

Sbjct: 1098 GGAEGETPHMLLRPHVEMPEVTPOMDYLPRVFN-QGIINPMLSDETNIEN 1147

Score = 198 (29.7 bits), Expect = 2.3e-11, Sum P(2) = 2.3e-11
Identities = 142/513 (27%), Positives = 200/513 (38%)

Query: 204 RPQSSPLLQPPAAQGTPEPCVQGHAAVRGLAFLPHQTVTIRFPCPVSLDAKQPCLLT 263
R + P +PP G + H V+ + +P L

Sbjct: 207 RTKKKPTPKPPVDEAGSGLDNGDFKVTTPDTSTTQHNVKSTSPKITTAKPINRPSLPP 266

Query: 264 R--TIRSTCLVHIEGDSVTKRVSARTNKARAP---ETPLSRRYDQAVTRPSR---AQTO 315
T + T L + +V+TK + TNK + E S + Q++ + S A T

Sbjct: 267 NSDTSKETSILTUNKETTVEKETT-TNKQSTSDGKEKTTSAKETQSIEKTSAKDLAPTS 325

Query: 316 GPVKAETPKAPFQICPGPMITKTLQTYPVVSVTLPQTYPASTMTTTPPKTSPVPKVITII 375
+ TPKA GP +T T + P T P+ PAST TP + +P +

Sbjct: 326 KVLAKPTPKAE-TTTKGPAIT-TPKEPTP---TTPKE-PAST---TPKEPTPTTIKSAP 375

Query: 376 KTPAQMYPGPTVTKTAPHTC--PMPTMTKIQVHPTASRTGTPRQTC-PATITAKNRQVVS 432
TP + P PT TK+AP T P PT TK + PT + P T PA T K+ P

Sbjct: 376 TTPKE--PAPTTTKSAPTTPEKAPPTTK-EPAPTTPEKAPTTTKEPAPTTTKSAPTT 432

Query: 433 ---LLASIMKSLPQVCPGPAKAPTPPMHPVTTPAKNPLQTCLSATMSKTSSQSPVGV 489
+ K P PA TP + P TTP K P T + T + +P

Sbjct: 433 KEPAPTTPKKPAPTTPEKAPTT-TPKEPTP-TTP-KEPAPTTKEPAPT-TPKEPAPTAPK 488

Query: 490 KPSPQT-RLPAMIT-KTPAQLRSVA---TILK---TLCLASPTVANVKAPPQVAVAGT 540
KP+P T + PA T K PA + T K T ++PT AP A T

Sbjct: 489 KPAPTTPEKAPTTPEKAPTTTKEPSPTTPEKAPTTTKSAPTTTKEPAPTTTKSAPTT 548

Query: 541 PNT-SGSIHENP----PKAKATVNVKQAAKV-KASSPSYLAEGKIRCLAQPHPGTGVPR 594
P S + + P PK A K+ A K +P+ E +P P P+

Sbjct: 549 PKEPSPTTKEPAPTTPEKAPTTPKKPAPTTPEKAPTTPEKAPTTTKEPAPTTTKEPAPTT 606

Query: 595 AAELPLEAEKIKTGQKQAKTDMAFKTSVAVEMAGAPSWTK-VAEEGDKPPHVYPVDM 653
A P ++ T K+ K + AP+ + +A + P P +

Sbjct: 607 EPA--PTTPKETAPTTPKKLTPTTPEKLAPTTPEKAPTTPEELAPTTPEEPTPTTPEEP 664

Query: 654 AVTLPRGQLAAPLTNASSQRHP-PCLSQRLAAPLTKASSQGHLPTELTKTPSLAHLDT 712
A T P+ AAP T + P P + P AP T P E T T

Sbjct: 665 APTTPKA--AAPNT----PKAPAPTTPEKAPTTTKEPSPTTPEKAPTTTKSAPTTTKEPAPTT 716

Query: 713 LSK 715
L +

Sbjct: 717 LKE 719

Score = 108 (16.2 bits), Expect = 4.3e-02, Sum P(2) = 4.3e-02
Identities = 60/214 (28%), Positives = 85/214 (39%)

Query: 265 TIRSTCLVHIEGDSVTKRVSAR-TNKA--RAPETP-LSRRYDQAVTRPSRAQTQGPVKA 320
T + +H D T +SA T KA +P+ P + A T+P T

Sbjct: 862 TTKEPTTIHKSPDE-STPELSAETPKALENSPKKEGVPVTTTKTPAATKPEMTTAKDKTT 920

Query: 321 ETP--KAPFQICPGPMITK-TLLQTYPVVSVTLPQTYPASTMTTTPPKTSPVPKVITIIKT 377
E P P +TK T T + T T TTT T+P K+T +KT

Sbjct: 921 ERDLRTTPETTTAAPKMTKETATTTTEKTTESKITATTTQVSTTTTQD-TTPF-KITTLKT 978

Query: 378 PAQMYPGPTVTK---TAPHTCPMPTMT-KIQVHPTASRTGTPRQTCPATITAKNRQVSL 433
+ P T TK T P T K + T S+ TP+ P A +P +

Sbjct: 979 TT-LAPKVTTTKKTIITTEIMNKPEETAKPKDRATNSKATTPKPQKPTK--APKKPTSTK 1035

Query: 434 LASIMKSL--PQVCPGPA-MAKTPPMHPVTTPAKNPLQT 470
M + P+ P P M T P+++P + A+ LQT

Sbjct: 1036 KPCTMPRVRKPKTTTPRKMTSTMPELNPTSRIAEAMLQT 1075

Score = 56 (8.4 bits), Expect = 3.1e-12, Sum P(2) = 3.1e-12

Identities = 17/60 (28%), Positives = 22/60 (36%)

Query: 22 TVHEPVVTQWAVHPPAPAHPSLLDKMEKAPPQPQHEGLKS-KEHLPQQPAEGKTASRRVP 80
T EP T P P PS E AP P+ + K+ P P E + + P
Sbjct: 533 TTKEPAPTTTKSAPTTTKEPSPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 592

Score = 52 (7.8 bits), Expect = 9.6e-16, Sum P(2) = 9.6e-16
Identities = 17/59 (28%), Positives = 22/59 (37%)

Query: 22 TVHEPV-VTQWAVHPPAPAHPSLLDKMEKAPPQPQHEGLKSKEHLPQQPAE-GKTASRR 78
T EP T P P P+ E P P+ +KE P P E TA ++
Sbjct: 431 TPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 489

Score = 51 (7.7 bits), Expect = 1.2e-15, Sum P(2) = 1.2e-15
Identities = 15/51 (29%), Positives = 19/51 (37%)

Query: 22 TVHEPVVTQWAVHPPAPAHPSLLDKMEKAPPQPQHEGLKS-KEHLPQQPAE 71
T EP T P P P+ + AP P+ + KE P P E
Sbjct: 416 TTKEPAPTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 466

Score = 47 (7.1 bits), Expect = 3.2e-15, Sum P(2) = 3.2e-15
Identities = 12/41 (29%), Positives = 17/41 (41%)

Query: 36 PPAHPSLLDKMEKAPPQPQHEGLKSKEHLPQQPAEGKTAS 76
P P P + P +P +KS P++PA T S
Sbjct: 350 PTPPTPK--EPASTTKEPTPTTIKSAPTTTKEPAPTTTTS 388

Score = 47 (7.1 bits), Expect = 3.2e-15, Sum P(2) = 3.2e-15
Identities = 15/57 (26%), Positives = 19/57 (33%)

Query: 22 TVHEPVVTQWAVHPPAPAHPSLLDKMEKAPPQPQHEG-LKSKEHLPQQPAEGKTASR 77
T EP T P P P+ E AP P+ +KE P T +
Sbjct: 377 TPKEPAPTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 433

Score = 46 (6.9 bits), Expect = 4.0e-15, Sum P(2) = 4.0e-15
Identities = 16/58 (27%), Positives = 22/58 (37%)

Query: 20 LATVHEPVVT---QWAVHPPAPAHPSLLDKMEKAPPQPQHEGLKSKEHLPQQPAEGKT 74
L T EP T + A P P+ + P +P KS P++PA T
Sbjct: 344 LTTKEPTPTTTPKEPASTTTPKEPTPTTIKSAPTTTKEPAPTTTTSAPTTTKEPAPTTT 401

Score = 42 (6.3 bits), Expect = 1.0e-14, Sum P(2) = 1.0e-14
Identities = 15/60 (25%), Positives = 21/60 (35%)

Query: 22 TVHEPVVTQWAVHPPAPAHPSLLDKMEKAPPQPQHEGLKS-KEHLPQQPAEGKTASRRVP 80
T EP T P P P+ + AP P+ + KE P E + + P
Sbjct: 463 TPKEPAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 522

Score = 39 (5.9 bits), Expect = 2.1e-14, Sum P(2) = 2.1e-14
Identities = 15/55 (27%), Positives = 20/55 (36%)

Query: 22 TVHEPVVTQWAVHPPAPAHPSLLDKMEKAPPQPQHEGLKSKEHLPQQPAEGKTAS 76
T EP T P PA + + P +P KS ++PA T S
Sbjct: 494 TPKEPAPTT---PKEPAPTTTKEPSPTTTPKEPAPTTTTSAPTTTKEPAPTTTTS 544

Pedant information for DKFZphtes3_4o19, frame 2

Report for DKFZphtes3_4o19.2

[LENGTH]	1180
[MW]	127693.40
[pI]	10.25
[HOMOL]	SWISSPROT:MUC2_HUMAN MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2). 1e-08
[FUNCAT]	98 classification not yet clear-cut [S. cerevisiae, YJR151c] 6e-06
[FUNCAT]	30.01 organization of cell wall [S. cerevisiae, YIR019c] 6e-06
[FUNCAT]	30.90 extracellular/secretion proteins [S. cerevisiae, YIR019c] 6e-06
[FUNCAT]	01.05.01 carbohydrate utilization [S. cerevisiae, YIR019c] 6e-06
[BLOCKS]	BL00412B Neuromodulin (GAP-43) proteins
[PROSITE]	CYTOCHROME_C 1
[PROSITE]	MYRISTYL 12
[PROSITE]	CAMP_PHOSPHO_SITE 1
[PROSITE]	CK2_PHOSPHO_SITE 8
[PROSITE]	PKC_PHOSPHO_SITE 25
[PROSITE]	ASN_GLYCOSYLATION 2
[KW]	Alpha Beta
[KW]	LOW_COMPLEXITY 5.00 %

SEQ MTLQGRADLSGNQGNAAGR LATVHEPVV TQWAVHP PAPAHP SLLDKMEKAP PQPQHEGLR
SEG
PRD cccccceeeccccccccceeeeeeeceeeeeeeccccccccceeecccccccccccccccc

SEQ SKEHLPQQPAEGKTASRRV PRLRAV VESQAFKN ILVDEMDMMHARAAT LIQANWRGYWLR
SEG
PRD cccccccccccccccccchhhhhhhhhhhhhhhheeehhhhhhhhhhhhhhhhhhccchhhh

SEQ QKLISQMAAKAIQEA WRRFRNKRHILHSSKSLVKKTRA EEGDIPYHAPQQVRFQHP EENR
SEG
PRD hhhhhhhhhhhhhhhhhhhhhhhheeeccccchhhhhhhhhccccccccceeeccccce

SEQ LLSPPI MVNKETQFPSCDN LVLCPQSSPLLQPPAAQGTPEPCVQGP HAAVRVGLAFLPH
SEG
PRD eccccceeeccccccccccccceeeccccccccccccccccccccccccccccceeeeeecccc

SEQ QTVTIRFP CPVSLDAKCPCLLTRTIRSTCLVHIEGDSVKTKRVSARTNKARAPETPLSR
SEG
PRD eeeeeccccccccccccccccccccceeeccccccccceeecccccccccccccccc

SEQ RYDQAVTRFSRAQTQGP VKAETPKAPFQICPGPMITKLLQTPVVS VTLPTQTPASTMT
SEG
PRD cccceeeccccccccceeecc

SEQ TTPPKTSPVPKVTI IKTPAQMPGPTVTKTAPHTCPMPTMTKI QVHPTASRTGT PRQTCP
SEG
PRD xxxxxxxxxxxxxxxx
ccccccccccccceeeccccccccccccccccccccccccccccceeecccccccccccccc

SEQ ATITAKNR PQVSLLASIMKSLPQVCPGPAMAKTPPQMHPVTTPAKNPLQTCLSATMSKTS
SEG
PRD ccc

SEQ SQRSVPVGTKPS PQTRL PAMITKTPAQLRSVATILKTLCLASPTVANVKAPQVAVAGT
SEG
PRD ccc

SEQ PNTSGSIHENPPKAKATVNVKQAAKVVKASSPSYLAEGKIRCLAQHPHPTGVPRAAAELP
SEG
PRD xxxxxxxxxxxxxxxx xxxxxx
cc

SEQ LEAEKIKTGTQKQAKTDMAFKTSVAVEMAGAPSWTKVAEEGDKPPHVYPVDMAVTLPRG
SEG
PRD xxxx
ccceeeccccceeecccccccccc

SEQ QLAAPLTNASSQRHP PCLSQRP LAAPLT KASSQGHLPTELTKT PSLAHLDTCLSKMHSQT
SEG
PRD ccc

SEQ HLATGAVKVQSQA PLATCLTKTQSRGQPIDITTC LIPAHQAADLSSNTHSQVLLTGSKV
SEG
PRD cccceeeccccceeeccccccccccccccccccccccccccccccccccccceeecccccc

SEQ SNHACQRLGGLSAPPWAKFEDRQTQPQPHGHVPGKTTQGGPCPAACEVQGMLVPPMAPTG
SEG
PRD ccc

SEQ HSTCNVESWGDNGATRAQPSMPGQAVPCQEDTG PADAGVVGQSWNRAWEPARGAASWDT
SEG
PRD ccc

SEQ WRNKAVVPPRRSGEPMVSMQAAEEIRILAVITIQAGVRGYLARRRIRLWHRGAMVIQATW
SEG
PRD cceeeccccccccchhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhh

SEQ RGYRVRRNL AHLCRATTTIQSAWRGYSTRDQARHWQMLHPVTWV ELSGRAGVMSDRSWF
SEG
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhheeeccccchhhhhhhhhhh

SEQ QDGRARTVSDHRCFQSCQAHACSVCHSLSSRIGSPPSVVMLVGSSPRTCHTCGRTOPTRV
SEG
PRD hccccceeeccccceeeccccceeeccccccccccccceeeccccccccccccccccce

SEQ VQGMGQGT EPGAVSWASAYQLAALS PRQPHRQDKAATAIQSAWRGFKIRQOMRQQQMAA
SEG
PRD eeccccccccccccchhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhh
xxxxxxxxxxxxxxxxxx

SEQ KIVQATWRGHHTR SCLKNTEALLGPADPSASSRHHMHWPGI
SEG
PRD xx
hhhhhhcccccehhhhhhhhhhcccccccccccccccccc

Prosites for DKFZphtes3_4o19.2

PS00001	542->546	ASN_GLYCOSYLATION	PDOC00001
PS00001	668->672	ASN_GLYCOSYLATION	PDOC00001
PS00004	282->286	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	76->79	PKC_PHOSPHO_SITE	PDOC00005
PS00005	148->151	PKC_PHOSPHO_SITE	PDOC00005
PS00005	244->247	PKC_PHOSPHO_SITE	PDOC00005
PS00005	265->268	PKC_PHOSPHO_SITE	PDOC00005
PS00005	278->281	PKC_PHOSPHO_SITE	PDOC00005
PS00005	281->284	PKC_PHOSPHO_SITE	PDOC00005
PS00005	285->288	PKC_PHOSPHO_SITE	PDOC00005
PS00005	288->291	PKC_PHOSPHO_SITE	PDOC00005
PS00005	299->302	PKC_PHOSPHO_SITE	PDOC00005
PS00005	322->325	PKC_PHOSPHO_SITE	PDOC00005
PS00005	414->417	PKC_PHOSPHO_SITE	PDOC00005
PS00005	424->427	PKC_PHOSPHO_SITE	PDOC00005
PS00005	481->484	PKC_PHOSPHO_SITE	PDOC00005
PS00005	610->613	PKC_PHOSPHO_SITE	PDOC00005
PS00005	671->674	PKC_PHOSPHO_SITE	PDOC00005
PS00005	679->682	PKC_PHOSPHO_SITE	PDOC00005
PS00005	900->903	PKC_PHOSPHO_SITE	PDOC00005
PS00005	959->962	PKC_PHOSPHO_SITE	PDOC00005
PS00005	987->990	PKC_PHOSPHO_SITE	PDOC00005
PS00005	1015->1018	PKC_PHOSPHO_SITE	PDOC00005
PS00005	1049->1052	PKC_PHOSPHO_SITE	PDOC00005
PS00005	1065->1068	PKC_PHOSPHO_SITE	PDOC00005
PS00005	1106->1109	PKC_PHOSPHO_SITE	PDOC00005
PS00005	1146->1149	PKC_PHOSPHO_SITE	PDOC00005
PS00005	1171->1174	PKC_PHOSPHO_SITE	PDOC00005
PS00006	22->26	CK2_PHOSPHO_SITE	PDOC00006
PS00006	42->46	CK2_PHOSPHO_SITE	PDOC00006
PS00006	156->160	CK2_PHOSPHO_SITE	PDOC00006
PS00006	546->550	CK2_PHOSPHO_SITE	PDOC00006
PS00006	848->852	CK2_PHOSPHO_SITE	PDOC00006
PS00006	988->992	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1003->1007	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1027->1031	CK2_PHOSPHO_SITE	PDOC00006
PS00008	11->17	MYRISTYL	PDOC00008
PS00008	14->20	MYRISTYL	PDOC00008
PS00008	539->545	MYRISTYL	PDOC00008
PS00008	591->597	MYRISTYL	PDOC00008
PS00008	746->752	MYRISTYL	PDOC00008
PS00008	777->783	MYRISTYL	PDOC00008
PS00008	853->859	MYRISTYL	PDOC00008
PS00008	878->884	MYRISTYL	PDOC00008
PS00008	882->888	MYRISTYL	PDOC00008
PS00008	1008->1014	MYRISTYL	PDOC00008
PS00008	1053->1059	MYRISTYL	PDOC00008
PS00008	1083->1089	MYRISTYL	PDOC00008
PS00190	1042->1048	CYTOCHROME_C	PDOC00169

(No Pfam data available for DKFZphtes3_4o19.2)

DKFZphtes3_50j4

group: testes derived

DKFZphtes3_50j4 encodes a novel 187 amino acid protein proline rich protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown, prolin ritch protein

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 1186 bp

Poly A stretch at pos. 1176, polyadenylation signal at pos. 1126

```

1  CACTGGGCGT CTGAAGCTCA GAGCTCACCC CTGAGATGGG CTCTCCTAGG
51  CCTCTGCGGA TGAGGGAGCC ACCAGGACCC AGTGTGTGA TGCTGTCTCT
101 TCCCTCTACC AGCACCTGCC CGCCAGAGA CCAGGGCACC CCTGAAGTCC
151 AGCCACCCCC TGCAAAGGAC ACATGGAAGG GCAAGCGGCC TCGATCCCAG
201 CAGGAGAACC CAGAGAGCCA GCCTCAGAAG AGGCCACGCC CCTCAGCCAA
251 GCCCTCCGTC GTAGCTGAGG TCAAGGGCAG CGTCTCGGCC AGCGAACAGG
301 GCACCTTGAA TCCCACGGCT CAAGACCCCT TCCAGCTCTC CGCTCCTGGC
351 GTCTCCTTGA AGGAGGCTGC AAATGTTGTG GTCAAGTGCC TCACCCCTTT
401 CTACAAGGAG GGCAAGTTTG CTTCCAAGGA GTTGTTTAAA GGCTTTGCCC
451 GCCACCTCTC ACACTTGCTG ACTCAGAAGA CCTCTCCTGG AAGGAGCGTG
501 AAAGAAGAGG CCCAGAACCT CATCAGGCAC TTCTTCCATG GCCGGGCCCG
551 GTGCGAGAGC GAAGCTGACT GGCATGGCCT GTGTGGCCCC CAGAGATGAC
601 CAACTGCTGG CTGGGCAGGG CCCGCGTCCT CCCCCAGATT CTAGCATGGG
651 TCATCCTGGG CCTCACCTGC TGATGCCAGG GCCATCGTCT TTTCTCAGTC
701 CTTCTCCTTT CCAACCATAC TTGGCTTTGG GGATGACCCC AGACACCCCC
751 TGAATCCAGG TCAGAGGTCA GCCCACCTTT CTTTCTGCTT GCAAAGCCTA
801 TAGACCCTTC TCAGAGCGGT CCTCATGGCT GGGTTTCTG GGACACATGT
851 CGAGGACAGA AGGTGGAGGG TGGTGGAGCT GCTGCTGGAA GAAGGGGAAG
901 GAAGAGTGGC CCTCCCCGA GTTCTAAGTC AGGATGAGGC CCACCTGTCC
951 AAGGTATCGG AACCTACCCA GGGGACCCTC AGATCCTCCA CCCACTCCCC
1001 CATCCATTAC GATGCCAGCT TCCAGCCTTG CCCAGGTGAG AGCTGTGGCA
1051 GAGGAGAGGC AGCCAGGCCG TGTTCCTGCT CAGCTCCTGC TCAGGAAGGC
1101 CAGGCCTGAC AGATGTTTGG GAGAGGAATA AAGTTGTGTT GTTGTGGGGC
1151 ATGCAGGCGT GCACACAGCC CTTTCAAAA AAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 36 bp to 596 bp; peptide length: 187
 Category: putative protein

```

1  MGSPRPFGMR EPPGPSAVMP ALPSTSTCPP RDQGTPEVQP TPAKDTWKVK
51  RPRSQQENPE SQPQKRPRPS AKPSVVAEVK GSVSASEQGT LNPTAQDPFQ
101 LSAPGVSLKE AANVVVKCLT PFYKEGKFAS KELFKGFARH LSHLLTQKTS
151 PGRSVKEEAQ NLIRHFFHGR ARCESEADWH GLCGPQR

```

BLASTP hits

Entry MMU92455_1 from database TREMBL:

product: "WW domain binding protein 7"; Mus musculus WW domain binding protein 7 mRNA, partial cds.
Score = 134, P = 6.9e-08, identities = 45/125, positives = 56/125

Alert BLASTP hits for DKFZphtes3_50j4, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_50j4, frame 3

Report for DKFZphtes3_50j4.3

```

[LENGTH]      187
[MW]           20353.06
[pI]           9.76
[PROSITE]      MYRISTYL      1
[PROSITE]      AMIDATION     1
[PROSITE]      CK2_PHOSPHO_SITE      6
[PROSITE]      PKC_PHOSPHO_SITE      6
[KW]           All Alpha
[KW]           LOW_COMPLEXITY      8.56 %

SEQ      MGSPPRPFGMREPPGPSAVMPALPSTSTCPPRDQGTPEVQPTPAKDTWKGKRPRSQQENPE
SEG      xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ      SQPQKRPRPSAKPSVVAEVKGSVSASEQGTLPNTAQDPFQLSAPGVSLKEAANVVVKCLT
SEG      .....
PRD      cccccccccccccchhhhhccccccccccccccccccccccccccccchhhhhhheeecc

SEQ      PFYKEGKFASKELFKGFARHLSHLLTQKTSPPGRSVKEEAQNLIHFFHGRARCESEADWH
SEG      .....
PRD      cccccccchhhhhhhhhhhhhhhhhheeeccccchhhhhhhhhhhhhhhccchhhhhhhhh

SEQ      GLCGPQR
SEG      .....
PRD      ccccccc

```

Prosites for DKFZphtes3_50j4.3

PS00005	3->6	PKC_PHOSPHO_SITE	PDOC00005
PS00005	46->49	PKC_PHOSPHO_SITE	PDOC00005
PS00005	70->73	PKC_PHOSPHO_SITE	PDOC00005
PS00005	107->110	PKC_PHOSPHO_SITE	PDOC00005
PS00005	146->149	PKC_PHOSPHO_SITE	PDOC00005
PS00005	154->157	PKC_PHOSPHO_SITE	PDOC00005
PS00006	54->58	CK2_PHOSPHO_SITE	PDOC00006
PS00006	84->88	CK2_PHOSPHO_SITE	PDOC00006
PS00006	94->98	CK2_PHOSPHO_SITE	PDOC00006
PS00006	107->111	CK2_PHOSPHO_SITE	PDOC00006
PS00006	154->158	CK2_PHOSPHO_SITE	PDOC00006
PS00006	175->179	CK2_PHOSPHO_SITE	PDOC00006
PS00008	81->87	MYRISTYL	PDOC00008
PS00009	48->52	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3_50j4.3)

DKFZphtes3_50n06

group: testes derived

DKFZphtes3_50n06 encodes a novel 186 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 1095 bp

Poly A stretch at pos. 1085, polyadenylation signal at pos. 1061

```
1 CAAGACCCTC GGAGCCAAGA AACAACTG AGTTCAGAT TTCGGAAGGT
51 TCACGAGTGT TGCCGACACG CCCTCCCAAC TGCAGACATC CTCCTGGAG
101 GACCTGCTGT GCTCACATGC CCCCTGTCC AGCGAGGACG ACACCTCCCC
151 GGGCTGTGCA GCGCCCTCCC AGGCACCCTT CAAGGCCTTC CTCAGTCCCC
201 CAGAGCCACA TAGCCACCGA GGCACCGACA GGAAGCTGTC CCCGCTCCTG
251 AGCCCTTGC AAGACTCACT GGTGGACAAG ACCCTGCTGG AGCCAGGGA
301 GATGGTCCGG CTAAGAAGG TGTGTTTCTC GGAGAGCAGC CTGCCACCG
351 GGGACAGGAC CAGGAGGAGC TACTACCTCA ATGAGATCCA GAGCTTCGCG
401 GGGCGCGAGA AGGACGCGCG CGTGGTGGGC GAGATCGCCT TCCAGCTGGA
451 CCGCCGCATC CTGGCCTACG TGTCCCGGG CGTGACGCGG CTCTACGGCT
501 TCACGGTGGC CAACATCCCC GAGAAGATCG AGCAGACCTC CACCAAGTCT
551 CTGGACGGCT CCGTGGACGA GAGGAAGCTG CCGGAGCTGA CGCAGCGTA
601 CCTGGCCCTG AGCGCGCGCC TGGAGAAGCT GGGCTACAGC CGCGACGTGC
651 ACCCGGGCTT CAGCGAGTTC CTCATCAACA CCTACGGAAT CCTGAAGCAG
701 CGGCGCGACC TGGCGGCCAA CCCCTGCAC AGCAGCCCGG CCGCGCTGCG
751 CAAGCTGGTC ATCGACGTGG TGCCCCCAA GTTCCTGGGC GACTCGCTGC
801 TGCTGCTCAA CTGCCTGTGC GAGCTCTCCA AGGAGGACGG CAAGCCCCCTC
851 TTCGCCTGGT GAGCCGCCCC GCGCCCGCG CTTGCCTGC AGTAAACGCG
901 TTTGTTCCAA CCCGGGGCCG CGGTGCCTCC TCGCGGTCCC CCCGGAGGGG
951 AAAGGGCCGC GTCCCCGCG CGCGAGGCCA GAGAAGGCC CGCTCCCACC
1001 GGTGCTGGGC CCGGACCGCA GCCCGCCGCT GCCCGCACCT GCGGAGTGCT
1051 TCTACCCCT CATTAAATC ATCCGTTTGC TTGTCAAAAA AAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 302 bp to 859 bp; peptide length: 186

Category: putative protein

Classification: no clue

```
1 MVRPKKVCFS ESSLPTGDRT RRSYYLNEIQ SFAGAEDAR VVGEIAFQLD
51 RRILAYVFPV VTRLYGFTVA NIPEKIEQTS TKSLDGSVDE RKLRELTQRY
101 LALSARLEKL GYSRDVHPAF SEFLINTYGI LKQRPDLRAN PLHSSPAALR
151 KLVIDVVPK FLGDSL LLLN CLCELSKEDG KPLFAW
```

BLASTP hits

No BLASTP hits available

No Alert BLASTP hits found

Report for DKFZphtes3 50n06.2

```
[LENGTH]      186
[MW]           21049.39
[pI]           9.28
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY      5.38 %
```

```
SEQ      MVRPKKVCFSSSLPTGDRTRRSYYLNEIQSFAGAEKDARVVGEIAFQLDRRILAYVFPG
SEG      .....
PRD      cccccceccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccc
```

SEQ VTRLYGFTVANIPEKIEQTSTKSLDGSVDERKLRELTQRYLALSARLEKLGYSRDVHPAF
SEG
PRD ceeeeeeeeeeccccccccccccccccchhhhhhhhhhhhhhhhhhhhhccccccccch

```
SEQ SEFLINTYGILKQRPDLRANPLHSSPAALRKLVIDVVPKFLGDSL LLLNCLCELSKEDG
SEG .....XXXXXXXXXXXXX.....
PRD hhhhhhccceeeccccccccccccccchhhhhhhhhhhcccccchhhhhhhhhhhccccc
```

```
SEQ      KPLFAW
SEG      . . . . .
PRD      CCCCCC
```

(No Prosite data available for DKFZphtes3 50n06.2)

(No Pfam data available for DKFZphtes3_50n06.2)

DKFZphtes3_50n23

group: testes derived

DKFZphtes3_50n23 encodes a novel 499 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

2 EST hits
(from other testis librarys) testis specific cDNA?

Sequenced by DKFZ

Locus: unknown

Insert length: 1907 bp

Poly A stretch at pos. 1897, polyadenylation signal at pos. 1872

```
1 GGGCACCAGC CACTTTCAC CATGACTGTG CGCTCGAGGG TCGCAGATGT
51 GTTCGGCAGC AAGGACACTG AGAGCCTTGA GCCTGTGCTT TTACCCTTAG
101 TAGATCGCAG GTTTCCTAAG AAATGGGAAA GACCGGTGGC AGAAAGCTTA
151 GGCCACAAAG ACAAGACCA GGAGGACTAC TTCCAGAAGG GAGGACTCCA
201 AATTAAAGTTC CACTGTAGCA AGCAGCTGTC TCTAGAGAGC TCCAGGCAGG
251 TGACCTCTGA GAGCCAAGAG GAGCCCTGGG AGGAGGAATT CGGCCGGGAG
301 ATGCCGGAGGC AGCTGTGGCT GGAGGAGGAG GAGATGTGGC AGCAGCGGCA
351 GAAGAAGTGG GCCCTGCTGG AGCAGGAGCA TCAGGAGAAG CTGCGGCAGT
401 GGAATCTGGA AGACCTGGCC AGGGAGCAAC AGCGGAGATG GGTCCAGCTA
451 GAAAAGGAGC AGGAGAGCCC ACGGAGAGAG CCAGAGCAGC TAGGGGAGGA
501 TGTGGAGAGG AGGATCTTCA CACCCACCAG TCGATGGAGG GACTTGGAGA
551 AGGCAGAGCT ATCATTAGTG CCTGCCCCAA GCCGGACCCA ATCTGCTCAC
601 CAAAGCAGGA GGCCACACTT GCCCATGTCT CCTAGTACCC AGCAGCCTGC
651 CCTGGGAAAG CAGAGACCTA TGAGTTCAGT GGAGTTTACC TACAGACCAC
701 GGACCCGCCG AGTTCACACA AAGCCCAAGA AATCTGCCTC CTTTCTGTGC
751 ACTGGGACAT CCATCCGAAG GCTGACCTGG CCCTCTTTGC AGATATCCCC
801 TGCAATATT AAGAAGAAGG TGTACCACAT GGACATGGAG GCCCAGAGGA
851 AGAACCTGCA GCTCCTGAGT GAGGAGTCTG AGTTGAGGCT GCCCACTAC
901 CTGCCGAGCA AAGCACTGGA GCTCACCACC ACCACCATGG AGCTGGGCGC
951 GCTCAGGCTG CAGTACCTGT GCCATAAGTA CATCTTCTAT AGACGCCTCC
1001 AGAGCCTCCG CCAAGAAGCG ATCAACCATG TACAAATCAT GAAAGAAACG
1051 GAGGCTTCCT ACAAGGCCCA GAACCTCTAC ATCTTCTGGA AAAACATTGA
1101 CCGCTGCAG AGTCTCAGGC TGCAGGCCTG GACGGACAAG CAGAAGGGGC
1151 TGGAGGAGAA GCACCGAGAG TGCCTGAGCA GCATGGTGAC CATGTTCCCC
1201 AAGCTCCAGC TGGAGTGGAA CGTTCACCTG AACATCCCTG AGGTCACCTC
1251 GCCAAGGCA AAGAAATGCA AGTTGCCTGC AGCCTCACCC CGGCACATCC
1301 GCCCAGTGG CCCCACCTAC AAGCAGCCCT TTCTGTCTAG GCACCGGGCA
1351 TGTGTGCCCC TGCAGATGGC CCGCCAACAG GGAAGCAGA TGGAGGCTGT
1401 CTGGAAGACC GAGGTGGCCT CCTCCAGTTA CGCAATAGAA AAAAGAGCC
1451 CTGCCAGCCT TCCCAGGAC CAGCTGAGGG GACACCCAGA TATTCCCAGG
1501 CTGTTGACAC TGGACGTGTA GTCCTCCTGC CACAAAAGCC TGAACCTCCT
1551 GAAGGCCAG TAAGCGCCTC AGCGAACCAG AGGAAGGAAT GCCAGGAACC
1601 TACAAATGAA TCCGCTTAGC TTGTTCAAAA AAAGTCAAGC GAGTCACTCC
1651 CTGGAACCCA AATAAGCCAG AAGGATCAAG ACAGCCCCAG TCTCCACTGC
1701 ATCCCTCAGC CAGTGATTCT CAACCTTCTG AGGGACGGA ACCCAGAG
1751 AACTTGGTCA AAATGCAGGT TCCCAGCTGG TGCTTTTAAA GAAACCTCT
1801 GGGGTTGCT GAGTACTCCT AGAACTTTGA GAAACACTGC TTCCCTCCTG
1851 CAGTCCCAAC ACTTACATT TTAATAAAAT AGAGGTTGGT TTATTTTAAA
1901 AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 22 bp to 1518 bp; peptide length: 499
 Category: similarity to known protein
 Classification: no clue

```

1  MTVRSRVADV FGSKDTESLE PVLLPLVDRR FPKKWERPVA ESLGHKDKDQ
51  EDYFQKGGGLQ IKFHCSKQLS LESSRQVTSE SQEEPWEEEF GREMRRQLWL
101 EEEEMWQQRQ KKWALLEQEH QEKLRQWNLE DLAREQQRW VQLEKEQESP
151 RREPEQLGED VERRIFTPTS RWRDLEKAEI SLVPAPSRTQ SAHQSRPHL
201 PMSPTQQA LGKQRPMSV EFTYRPRTRR VPTKPKKSAS FVVTGTSIRR
251 LTWPSLQISP ANIKKKVYHM DMEAQKRLQ LLSESELRL PHYLRKALE
301 LTTTLMELGA LRLQYLCHKY IFYRRLQSLR QEAINHVQIM KETEASYKAQ
351 NLYIFLENID RLQSLRLQAW TDKQKGLEEK HRECLSSMT MFPKLQLEWN
401 VHLNIPVETS PKPKCKLPA ASPRHIRPSG PTYKQPFSLR HRACVPLQMA
451 RQQKQMEAV WKTEVASSY AIEKTPASL PRDQLRGHPD IPRLLTLDV

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_50n23, frame 1

PIR:S28589 trichohyalin - rabbit, N = 1, Score = 134, P = 5.3e-05

TREMBLNEW:AF132479_1 product: "Ese2L protein"; Mus musculus Ese2L
 protein mRNA, complete cds., N = 1, Score = 130, P = 0.00017

>PIR:S28589 trichohyalin - rabbit
 Length = 1,407

HSPs:

Score = 134 (20.1 bits), Expect = 5.3e-05, P = 5.3e-05
 Identities = 88/354 (24%), Positives = 154/354 (43%)

Query: 29 RRFPPKKWERPVAESLGHKDKDQEDYFQKGGGLQIK-FHCSKQLSLESSRQVTSESQEEPWE 87
 R++ K +R + L + ++E ++ G + F +QL +++ E +EE +
 Sbjct: 165 RQYRDKEQLRQLEERRAEELRRRGRDAEEFIEEQLRRREQQELKRELREEEQ 224

Query: 88 EEFGREMRRQLWLEEEEMWQQRQKKWALLEQEHQEKLRQWNLEDLAREQQRWVQLEKEQ 147
 RE + L+EEE RQ++W E Q++LR+ LE++ RE+++R Q E+ +
 Sbjct: 225 RRRERREQHERA-LQEEELQLRQRWRE-EPREQQQLRR-ELEEI-REREQRLEQEERRE 280

Query: 148 ESPRREPEQLGEDVERRIFTPTSRWRDLEKAELSVPAPSRTQSAHQSRPHLPMSPSTQ 207
 + RRE ++L E ERR ++ + E L R Q Q R + +
 Sbjct: 281 QQLRRE-QRL-EQEERREQQLRRELEIREREQRLEQEERREQLEQEERREQQLKRELE 338

Query: 208 QPALGKQRPMSVVEFTYRPRTRRVPTKPKKSASFVVTGTSIRRLTWPSLQISPANIKK-K 266
 + +QR +E R R + + + ++ A G S + R W S A ++ K
 Sbjct: 339 EIREREQR---LEQEER-REQLLAEEVREQAR--ERGSLTR-RWQRQLESEAGARQSK 390

Query: 267 VYHMDMEAQKRLQLLSESELRLPHYLRKALELTTTMM-----ELGALRLQYLCHKY 320
 VY +R+ Q L ++ E R R + LE E R Q L +
 Sbjct: 391 VYS---RPRRQEEQSLRQDQERR-QQERERELEEQAARRQQWQAEESERRRQRLSARP 446

Query: 321 IFYRRLQSLRQEAINHVQIMKETEASYKAQONLYI-FLENIDRLQSL-RLQAWTDKQKGLE 378
 R Q +E Q +E E + + + FLE ++LQ R Q ++ E
 Sbjct: 447 SLRER-QLRAEERQEQRFRREEEQRRRRQELQFLEEEELQRRERAAQLQEEDSFQE 505

Query: 379 EKHR 382
 ++ R
 Sbjct: 506 DRER 509

Score = 119 (17.9 bits), Expect = 2.2e-03, P = 2.2e-03
 Identities = 79/357 (22%), Positives = 150/357 (42%)

Query: 33 KWERPVAESLGHKDKDQEDYFQKGGGLQIKFHCSKQLSLESSRQVTSESQEEPWEEFGR 92
 ++ E+ + + K +++E Q+ + + +Q R+ + + + EE+F +
 Sbjct: 990 RREEQLRQERDRKFRREEQLQE---REERLRQERDRKFRREEERQLRRQLEEQFRQ 1046

Query: 93 EMRRQLWLEEEEMWQQRQKKWALLEQEHQEKLRQWNLEDLAREQQRWVQLEKEQESPRR 152
 E R+ LEE+ + Q+++K L QE K R+ E+ R +Q R QL +E++ R
 Sbjct: 1047 ERDRKFRLEEQ-IRQEKEEK-QLRRQERDRKFR---EEQRRRQEREQQLRRERDRKFR 1101

Query: 153 EPEQLGEDVERRIFTPTSRWRDLEKAELSVPAPSRTQSAHQSR--RPHLPMSPSTQQA 210

Sbjct: 1102 E EQL ++ E R R L + E L + + + R R + +++
EEEEQLQEREERLRQERARKLREEE-QLLRREEQLLRQERDRKFREEEQLLQESEER 1160

Query: 211 LGKQ---RPMSSVEFTYRPTRRVPTKPKKSASFVPTGTSIRRLTWPSLQISPAKIKKV 267
L + Q R + E + R + +++ +R+ Q ++++

Sbjct: 1161 LRRQERERKLREEEQLLQEREERLRQERARKLREEEQLLRQEEQLRQERARKLREEE 1220

Query: 268 YHMDMEAQ-----RKNLQLLS-ESELRLPHYLRSKALELTTTMMELGALRLQYL 316
+ E Q R+ QLL EE ELR + + E E LR Q

Sbjct: 1221 QLLRQEEQLRQERDRKFREEEQLLRREEQLRRERDRKFREEEQLLQEREERLRQER 1280

Query: 317 CHKYIFYRRLQSLRQEAINHVMQIMKETEASYKAQNLIFYFLENIDRLQ-SLRLQAWTDKQK 375
K + L E ++ +E + Y+A+ + E RL+ LR + +++

Sbjct: 1281 ARK--LREEEQLLFEEQEEQLRQERDRRYRAEEQFAREEKSRLRELRLQEEEQRRR 1338

Query: 376 GLEEKHRE 383
E K RE

Sbjct: 1339 ERERKFRE 1346

Score = 109 (16.4 bits), Expect = 1.9e-01, P = 1.7e-01
Identities = 37/113 (32%), Positives = 60/113 (53%)

Query: 67 QOLSLESSRQVTSSESQ--EEPWEEFGRMRRQLWLEEEEMWQQRQKKWALLEQEHQEK 124
+QL E R+ E Q +E EE R+ R + EEE++ Q+R+++ L QE + KL

Sbjct: 764 QLLRRERDRKFREEEQLLQEREERLRQERERKLREEEQLLQEREER-RLRRQERERKL 822

Query: 125 RQWNLEDLAREQRRWVQLEKEQESPRREPEQLGEDVERRIFTPTSRWRDLEKAE 179
R+ E L +E++ ++ +E+E RE EQL E + + R R L + E

Sbjct: 823 REE--EQLQEREERLR-RQERERKLREEEQLLRQEEQEL--RQERARKLREEE 872

Score = 107 (16.1 bits), Expect = 3.0e-01, P = 2.6e-01
Identities = 35/109 (32%), Positives = 61/109 (55%)

Query: 71 LESSRQVTSSESQEEPWE-EFGRMRRQL---WLEEEEMWQQRQKKWALLEQEHQEKLRQ 126
L Q+ ES+EE +E +++RR+ + EEE++ Q+R+++ L QE + KLR+

Sbjct: 742 LREEEQLLQEEERLRQEREQQLRRERDRKFREEEQLLQEREER-RLRRQERERKLRE 800

Query: 127 WNLEDLAREQRRWVQLEKEQESPRREPEQLGEDVERRIFTPTSRWRDLEKAE 179
E L +E++ ++ +E+E RE EQL ++ E R R L + E

Sbjct: 801 E--EQLQEREERLR-RQERERKLREEEQLLQEREERLRQERERKLREEE 850

Score = 104 (15.6 bits), Expect = 9.4e-02, P = 9.0e-02
Identities = 84/339 (24%), Positives = 149/339 (43%)

Query: 67 QOLSLESSRQVTSSESQEEPWEEFGRMRRQL-WLEEEEMWQQRQKKWALLEQE--HQEK 123
+QL E ++ +EE EE RE R++L +LEEE Q+R++ L E++ +++

Sbjct: 451 QLRAEEERQEQRFREE---EQRRERRRQELQFLEEEQQLRRERAQQLQEEDSFQEDR 507

Query: 124 LRQWNLEDLAREQRRWVQLEKEQESPRR---EP---EQLGEDVE-RRIFTPTSRWRDL 175
R+ ++ Q RW QL++E + R +P EQL E+ E +R R R+

Sbjct: 508 ERRRRQQEQRPQGTWRW-QLQEEAQRRTLYAKPGQQEQRLREEELQREKRRQEREREY 566

Query: 176 EKAELSLVPAPSRTQSAHQSRPHLPMPSTQPPALGKQRPMSSEFTYRPT----RRV 231
+ E L + + R + + Q+ L + R + E + R RR

Sbjct: 567 REEE-KLQREDEKRRRQERERQYRELEELRQEEQL-RDRKLREEEQLLQEREERLRQ 624

Query: 232 PTKPK---KSASFVPTGTSIRRLTWPSLQISPAKIKKVYHMDMEAQRK---NLQLLSEE 285
+ K + +R+ L+ +++ + E +RK QLL E

Sbjct: 625 ERERKLREEEQLLRQEEQLRQERERKLREEEQLLRREEQLRQERERKLREEEQLLQER 684

Query: 286 SELRLPHYLRSKALE-----LTTTMMELGALRLQYLCHKYIFYRRL-QSLRQEAINHVV-- 337
E RL R++ L L EL R + L + RR Q LRQE +

Sbjct: 685 EEERLRQERARKLREEEQLLRQEEQLRQERERKLREEEQLLRREEQLLRQERDRKLRE 744

Query: 338 --QIMKETEASYKAQNLIFYFLENIDRLQSLRLQAWTDKQGLEEKHRECL 385
Q+++E+E + E +L+ R + + +++ L+E+ E L

Sbjct: 745 EEQLQEESEERLRQ-----EREQQLRRERDRKFREEEQLLQEREERL 789

Score = 103 (15.5 bits), Expect = 1.2e-01, P = 1.1e-01
Identities = 42/152 (27%), Positives = 74/152 (48%)

Query: 36 ERPVAESLGHKDKQEDYFQKGLQIKFHCSKOLSLESSRQVTSSESQEEPWEEFGR-REM 94
ER + K +++E ++ +++ +++L E + + E QE E + RE

Sbjct: 835 ERLRRQERERKLREEEQLLRQEEQLRQERARKLR-EEEQLLRQEEQLRQERDRKLREE 893

Query: 95 RRQLWLEEEEMWQQRQKKWA----LLEQEHQEKLRQWNLEDLAREQ---RRWVQ-LEKE 146
+ L EE+E+ Q+R +K LL++ +E+LR+ E RE++ RR Q L +E

Sbjct: 894 EQLLRQEEQLRQERDRKLREEEQLLQESEERLRQERERKLREEEQLLRREEQELRRE 953

Query: 147 QESPRREPEQLGEDVERRIFTPTSRWRDLEKAE 179
+ RE EQL ++ E R R L + E

Sbjct: 954 RARKLREEEQLLQEREERLRRQERARKLREEE 986

Score = 103 (15.5 bits), Expect = 7.8e-01, P = 5.4e-01
Identities = 31/91 (34%), Positives = 52/91 (57%)

Query: 67 KQLSLESSRQVTSSESQEEPWEEFGRMRRQLWLEEEEMWQQRQKKWALLEQEHQEKLRQ 126
++L E R++ E Q EE+ R+ R+ EEE++ Q+R+++ L QE KLR+

Sbjct: 642 QELRQERERKLREEEQLLRREEQELRQERERKLREEEQLLQEREERLRRQERARKLRE 700

Query: 127 WNLEDLAREQRRWVQLEKEQESPRREPEQL 157
E L R++++ +L +E+E RE EQL

Sbjct: 701 E--EQLLRQEEQ---ELRQERERKLREEEQ 726

Score = 101 (15.2 bits), Expect = 2.0e-01, P = 1.8e-01
Identities = 38/111 (34%), Positives = 57/111 (51%)

Query: 72 ESSRQVTSSESQEEPWEEFGRMRRQLWLEEEEMWQQRQKKWALLEQEHQEKLRQWNLE 130
E R++ E Q EE E RE R+L EEE++ Q+R+++ L QE KLR+ +

Sbjct: 931 ERERKLREEEQLLRREEQELRRERARKL-REEEQLLQEREERLRRQERARKLREEE-Q 987

Query: 131 DLAREQRRWVQLEKEQESPRREPEQLGEDVERRIFTPTSRWRDLEKAELS 182
L RE+Q +L +E++ RE EQL ++ E R R + E L

Sbjct: 988 LLRREEQ---ELRQERDRKFREEEQLLQEREERLRRQERDRKFREERQ 1035

Score = 101 (15.2 bits), Expect = 1.3e+00, P = 7.2e-01
Identities = 33/108 (30%), Positives = 56/108 (51%)

Query: 72 ESSRQVTSSESQEEPWEEFGRMRRQLWLEEEEMWQQRQKKWALLEQEHQEKLRQWNLE 131
E R++ E Q EE+ R+ R+ EEE++ +Q +++ L QE KLR+ E

Sbjct: 841 ERERKLREEEQLLRQEEQELRQERARKLREEEQLLRQEEQ---LRQERDRKLREE--EQ 895

Query: 132 LAREQRRWVQLEKEQESPRREPEQLGEDVERRIFTPTSRWRDLEKAE 179
L R++++ +L +E++ RE EQL ++ E R R L + E

Sbjct: 896 LLRQEEQ---ELRQERDRKLREEEQLLQESEERLRRQERERKLREE 940

Score = 99 (14.9 bits), Expect = 2.0e+00, P = 8.7e-01
Identities = 32/97 (32%), Positives = 50/97 (51%)

Query: 72 ESSRQVTSSESQEEPWEEFGRMRRQLWLEEEEMWQQRQKKWALLEQEHQEKLRQWNLE 131
E R+ E Q EE E R L EEE Q +++ L QE + KLR+ E

Sbjct: 578 EKRRRQERERQYRELELRQEEQLRDRKLREEEQLLQEREERLRRQERERKLREE--EQ 635

Query: 132 LAREQ----QRRWVQLEKEQESPRREPEQLGEDVERRI 165
L R++ Q R +L +E++ RRE ++L ++ ER++

Sbjct: 636 LLRQEEQELRQERERKLREEEQLLRREEQELRQERERKL 674

Score = 99 (14.9 bits), Expect = 2.0e+00, P = 8.7e-01
Identities = 34/111 (30%), Positives = 58/111 (52%)

Query: 67 KQLSLESSRQVTSSESQ--EEPWEEFGRMRRQLWLEEEEMWQQRQKKWALLEQEHQEK 124
++L E R++ E Q +E EE R+ R+ EEE++ +Q +++ L QE + KL

Sbjct: 664 QELRQERERKLREEEQLLQEREERLRRQERARKLREEEQLLRQEEQ---LRQERERKL 720

Query: 125 RQWNLEDLAREQRRWVQLEKEQESPRREPEQLGEDVERRIFTPTSRWRDLEK 177
R+ + L RE+Q L +E++ RE EQL ++ E R + L +

Sbjct: 721 REEE-QLLRREEQ---LRQERDRKLREEEQLLQESEERLRRQEREQQLRR 768

Score = 98 (14.7 bits), Expect = 2.6e+00, P = 9.2e-01
Identities = 37/146 (25%), Positives = 77/146 (52%)

Query: 20 EPVLLPLVDRFPKWERPVAESLGHKDKQEDYFQKGLQIKFHCQKQLSLESSRQVTS 79
E LL ++ ++ ER + E + +E+ ++ K +QL + +++

Sbjct: 655 EEQLLRREEQELRQERERKLREEEQLLQEREERLRRQERARKLREEEQLLRQEEQELRQ 714

Query: 80 ESQEEPWEEFGRMRRQLWLEEEEMWQQRQKKWALLEQEHQEKLRQWNLE-LAREQQR 138
E + + EEE + +RR+ L +E ++ +++ LL++ +E+LR+ E L RE+ R

Sbjct: 715 ERERKLREEE--QLLRREEQLLRQERDRKLREEEQLLQESEERLRRQEREQQLRRERDR 772

Query: 139 RWVQLEKEQESPRREPEQLG-EDVERRI 165
++ E+EQ RE E+L ++ ER++

Sbjct: 773 KF--REEEQLLQEREERLRRQERERKL 798

Score = 97 (14.6 bits), Expect = 3.3e+00, P = 9.6e-01
Identities = 38/129 (29%), Positives = 63/129 (48%)

Query: 72 ESSRQVTSSESQ--EEPWEEFGRMRRQLWLEEEEMWQQRQKKWALLEQEHQEKLRQWN 129
E R++ E Q +E EE R+ R+ EEE++ +Q +++ L QE KLR+

Sbjct: 817 ERERKLREEEQLLQEREERLRRQERERKLREEEQLLRQEEQ---LRQERARKLREE-- 871

Query: 130 EDLAREQRRWVQLEKEQESPRREPEQLGEDVERRIFTPTSRWRDLEKAELSLVPAPSRT 189
E L R++++ +L +E++ RE EQL E+ + R R L + E L+

Sbjct: 872 EQLLRQEEQ---ELRQERDRKLREEEQLLRQEEQEL--RQERDRKLREEE-QLLQESEEE 925

Query: 190 QSAHQSRPHL 200

+ Q R L

Sbjct: 926 RLRRQERERKL 936

Score = 96 (14.4 bits), Expect = 4.1e+00, P = 9.8e-01
Identities = 41/132 (31%), Positives = 69/132 (52%)

Query: 46 KDKDQEDYFQKGGGLQI-KFHCSKQLSLESSRQVTSESQEEPWEEFGRMRRQLWLEEEE 104
+++ QE F + Q + ++QL E S Q E + E + G + R QL +EE

Sbjct: 473 RERRQELQFLEEEQLQRRERAQQLQEEDSFQEDRERRRRQEQRPQTWRWQL---QEE 529

Query: 105 MWQQRQKKWALLEQEHQEKLRQWNLEDLAREQRRWVQLEKEQESPRPEQLGEDVERR 164

++R +A Q QE+LR+ E+L RE++R+ E+E+E E Q ED +RR

Sbjct: 530 AQRRRHTLYAKPGQ--QEQLREE--EELQREKRQ----EREREYREEEKLQREDEKRR 581

Query: 165 IFTPTSRWRDLEK 177

++R+LE+

Sbjct: 582 RQERERQYRELEE 594

Score = 96 (14.4 bits), Expect = 4.1e+00, P = 9.8e-01
Identities = 35/138 (25%), Positives = 76/138 (55%)

Query: 28 DRFPKKWERPVAESL-GHKDKDQEDYFQKGGGLQIKFHCSKQLSLESSRQVTSESQEEPW 86

+R++ + E E L K ++E Q + ++ L Q + ++E

Sbjct: 586 EQYRELEELRQEEQLRDRKLREEEQLLQEREERLRRQERERKLREEEQLLRQEEQE-L 644

Query: 87 EEEFGRMRRQLWL---EEEMWQQRQKKWALLEQEHQEKLRQWNLEDLAREQRRWVQL 143

+E R++R + L EE+E+ Q+R++K L +E Q L++ E L R++ R +L

Sbjct: 645 RQERERKLREEEQLLRREEQELRQERERK---LREEEQ-LLQEREERLRRQERAR--KL 698

Query: 144 EKEQESPRPEQLGEDVERRI 165

+E++ R+E ++L ++ ER++

Sbjct: 699 REEEQLLRQEEQELRQERERKL 720

Score = 95 (14.3 bits), Expect = 5.2e+00, P = 9.9e-01
Identities = 59/282 (20%), Positives = 121/282 (42%)

Query: 20 EPVLLPLVDRFPKKWERPVAESLGHKDKDQEDYFQKGGGLQIKFHCSKQLSLESSRQVTS 79

E LL ++ ++ ER + E + +E+ ++ K +QL + ++

Sbjct: 655 EEQLLRREEQELRQERERKLREEEQLLQEREERLRRQERARKLREEEQLLRQEEQELRQ 714

Query: 80 ESQEEPWEEFGRMRRQLWLEEEEMWQQRQKKWALLEQEHQEKLRQWNLED-LAREQQR 138

E + + EEE + +RR+ L +E + + ++ LL++ +E+LR+ E L RE+ R

Sbjct: 715 ERERKLREEE--QLLRREEQLLRQERDRKLREEEQLLQESEEERLRRQEREQQLRRERDR 772

Query: 139 RWVQLEKEQESPRPEQLG-EDVERRIFTPTSRWRDLEKAELS LVPAPSRQTSAHQ--S 195

++ E+EQ RE E+L ++ ER++ ++ E+ L + + Q

Sbjct: 773 KF--REEEQLLQEREERLRRQERERKLREEEQLLQEREERLRRQERERKLREEEQLLQ 830

Query: 196 RRPPLPMSPTQPALGKQRPMSVEFTYRPRTRRVPTKPKKSASFVVTGTSIRRLTWPS 255

R + ++ L ++ + E R R ++ +R+

Sbjct: 831 EREEERLRRQERERKLREEEQLLRQE-EQELRQERARKLREEEQLLRQEEQELRQERDRK 889

Query: 256 LQISPANIKKKVYHMDMEAQRK---NLQLLSESELRLPHYLRSKAL 299

L+ +++ + E RK QLL E E RL R + L

Sbjct: 890 LREEEQLLRQEEQELRQERDRKLREEEQLLQESEEERLRRQERERKL 936

Score = 94 (14.1 bits), Expect = 1.1e+00, P = 6.8e-01
Identities = 35/116 (30%), Positives = 59/116 (50%)

Query: 72 ESSRQVTSESQEEPWEEFGRMRRQLWLEEEEMWQQRQKKWALLEQEHQEK-----L 124

E +R++ E Q EE+ R+ R + + EEE++ Q+R+++ L QE K L

Sbjct: 977 ERARKLREEEQLLRREEQELRQERDRKFREEEQLLQEREEE-RLRRQERDRKFREEERQL 1035

Query: 125 RQWNLEDLAREQRRWVQLEKEQESPRPEQLGEDVERRIFTPTSRWRDLEKAELS 182

R+ LE+ R+++ R +LE EQ +E +QL R F + R ++ E L

Sbjct: 1036 RRQELEEQFRQERDRKFRLE-EQIRQEKEEKQLRRQERDRKFREEEQRRRQEREQQL 1092

Score = 94 (14.1 bits), Expect = 1.1e+00, P = 6.8e-01
Identities = 51/166 (30%), Positives = 76/166 (45%)

Query: 67 QLSLESSRQVTSESQ--EEPWEEFGRMRRQLWLEEEEMWQQRQKKWALLEQEHQEK 123

++L E R+ E Q +E EE R+ R R+L EEE++ + Q++ L QE+

Sbjct: 1250 QELRRERDRKFREEEQLLQEREERLRRQERARKLREEEQLLFEEQEEQRL----RQER 1305

Query: 124 LRQWNLED-LAREQRRWVQLEKEQESPRPEQLGEDVERRIFTPTSRWRDLEKAELS 182

R++ E+ ARE++ R +LE+E R+E EQ R F R E+ E

Sbjct: 1306 DRRYRAEEQFAREEKSR--RLEREL---RQEEQRRRRERERKFREEQLRRQEE-EQRR 1359

Query: 183 VPAPSRQTSAHQSRPHLPMSPTQQPALGKQRPMSVVEFTYRPRTRRVP 232
 R QSR L P T+Q A R E+ R++ P
 Sbjct: 1360 RQLRRERQFREDQSRQVL--EPGTRQFARVPVRSSPLYEYIQEQRSYRP 1407

Score = 93 (14.0 bits), Expect = 8.3e+00, P = 1.0e+00
 Identities = 41/145 (28%), Positives = 72/145 (49%)

Query: 28 DRRFPKKWERPVAESLGHKDKDQEDYFQKGGLOIKFHCSKQLSLESSRQVTSESQEEPW- 86
 +RR ++ ER + E ++ Q + + Q + L R + QE+ +
 Sbjct: 408 ERRQRQERERELEEQARRQQWQAEESERRRQ-RLSARPSLRERQLRAEERQEERQFR 466

Query: 87 -EEEFGRMRRQL-WLEEEEMWQQRQKKWALLEQE--HQEKLQWNLEDLAREQRRWVQ 142
 EEE RE R++L +LEEEE Q+R++ L E++ +++ R+ ++ Q RW Q
 Sbjct: 467 EEEQRERRRQELQFLEEEELQRRERAAQLQEEDSFQEDRERRRRQEQRPQTWRW-Q 525

Query: 143 LEKEQESPRR----EP---EQLGEDVE 162
 L++E + R +P EQL E+ E
 Sbjct: 526 LQEEAQRRTLYAKPGQEQLEEEEE 552

Score = 91 (13.7 bits), Expect = 2.4e+00, P = 9.1e-01
 Identities = 38/110 (34%), Positives = 57/110 (51%)

Query: 72 ESSRQVTSESQEEPWEE-EFGREMRRQLWLEEEEMWQQRQKKWALLEQEHQEKLRQWNL- 129
 E R++ E Q EE E RE R+L EEE++ Q+R+++ L QE KLR+
 Sbjct: 931 ERERKLREEEQLLRREEQELRRERARKL-REEEQLLQEREEE-RLRRQERARKLREEEQ 988

Query: 130 -----EDLAREQRRWVQLEKEQESPRPEQLGEDVERRIFTPTSRWRDLEKAEL 180
 ++L +E+ R++ E+EQ RE E+L R F R L + EL
 Sbjct: 989 LRREEQELRQERDRKF--REEEQLLQEREEELRRQERDRKFREEER--QLRRQEL 1040

Score = 89 (13.4 bits), Expect = 2.2e+00, P = 8.9e-01
 Identities = 35/138 (25%), Positives = 65/138 (47%)

Query: 82 QEEPWEEFGRMRRQLWLEEEEM--WQQRQKKWALLEQEHQEKLRQWNLEDLAREQRR 139
 Q E++ E+R + + +E E WQ++++ L E+E Q K R+ + +R+ + +
 Sbjct: 111 QNRQEDQRRFELRDRQFEDEPERRRWQKQEQERELAEEEQRRKRERFEQHYSRQYRDK 170

Query: 140 WVQLEKEQ-ESPRPEQL---GEDVERRIFTPTSRWRDLEKAELSLVPAPSRQTSAHQ 194
 +L++++ E R E EQL G D E F + R E+ EL Q +
 Sbjct: 171 EQRLRQLEERAEELRRRRGRDAEE--FIEEEQLRRREQELKR-ELREEEQRRRE 227

Query: 195 SRRPHLPMSPTQQPALGKQR 215
 R H ++ L ++R
 Sbjct: 228 RREQHERALQEEELQLRQR 248

Score = 50 (7.5 bits), Expect = 2.2e+00, P = 8.9e-01
 Identities = 34/160 (21%), Positives = 67/160 (41%)

Query: 325 RLQSLRQEAINHVMQIMKETEASYKAQNLIFLENIDRL-QSLRLQAWTDKQGLEEKHRE 383
 R + R+E Q+ +E E + + LE +R Q LR + +++ E++ R
 Sbjct: 245 RQRRWREEPREQQLRRELEIREREQR---LEQERREQLRREQRLEQERREQLRR 301

Query: 384 CLSSMVTMFPKLQLEWNVHLNIP-EVTSPPKPKCKLPAASPRHIRPSGPTYKQPFSLRHR 442
 L + +L+ E + E + K+L R R ++ L+
 Sbjct: 302 ELEEIREREQRLEQERREQRLEQERREQLKRELEEIREREQRLEQERREQLLAEV 361

Query: 443 ACVPLQMARQGGKQMEAVWKTEVASSYAIEKKTASLPRDQ 484
 + AR++G+ + W+ ++ S + A + K S PR Q
 Sbjct: 362 R----EQAREGESLRRWQRQLESEAGARQSKV-YSRPRRQ 398

Score = 40 (6.0 bits), Expect = 1.9e-01, P = 1.7e-01
 Identities = 32/115 (27%), Positives = 47/115 (40%)

Query: 276 RKNLQLLSEESLRLPHYLRSKAL--ELTTTMMELGALRLQYLCHKYIFYRRL-QSLRQE 332
 R+ QLL E E RL R++ L E E LR Q K+ +L Q +E
 Sbjct: 959 REEQQLQEREEELRRQERARKLREEEQLLRREEQELR-QERDRKFREEEQLLQEREEE 1017

Query: 333 AINHVMQIMKETEASYKAQNLIFLENIDRLQSLRLQAWTDKQ-KGLEEKHRE 383
 + + +E E + Q L F + DR L Q +K+ K L + R+
 Sbjct: 1018 LLRRQERDRKFREEERQLRQEEELQFRQERDRKFRLEEQIRQEKEEKQLRRQERD 1073

Score = 37 (5.6 bits), Expect = 1.6e+00, P = 7.9e-01
 Identities = 27/108 (25%), Positives = 43/108 (39%)

Query: 276 RKNLQLLSEESLRLPHYLRSKAL--ELTTTMMELGALRLQYLCHKYIFYRRLQSLRQE 332
 R+ QLL E E RL R + L E E LR Q K R + L QE
 Sbjct: 775 REEQQLQEREEELRRQERERKLREEEQLLQEREEELRRQERERKL---REEEQLLQE 831

Query: 333 AINHVMQIMKETEASYKAQNLIFLENIDRLQSLRLQAWTDKQGLEEKHRE 383
 +E E + + + E L+ R + +++ L ++ +E
 Sbjct: 832 REEELRRQERERKLREEEQLLRQEE-QELRQERARKLREEEQLLRQEEQE 881

Pedant information for DKFZphtes3_50n23, frame 1

Report for DKFZphtes3_50n23.1

```
[LENGTH]      499
[MW]           58885.69
[pI]           9.67
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY      10.42 %

SEQ  MTVRSRVADVFGSKDTESLEPVLLPLVDRRFPPKKWERPVAESLGHKDKDQEDYFQKGGLQ
SEG  .....
PRD  cccccceeecccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  IKFHCSKQLSLESSRQVTSESQEEPWEEEFGREMRRQLWLEEEEMWQQRQKKWALLEQEH
SEG  .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
PRD  eeeecchhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  QEKLRQWNLEDLAREQQRRWVQLEKEQESPREPEQLGEDVERRIFTPTSRRWRDLEKAEL
SEG  .....
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ  SLVPAPSRQTSAHQSRPHLPMSPTQQPALGKQRPMSVVEFTYRPRTRRVPTKPKKSAS
SEG  .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
PRD  hccccccchhhhhccccccccccccccccccccccccccccccccccccccccccccccccccce

SEQ  FPVGTGSIRRLTWPSLQISPANIKKKVYHMDMEAQRKNLQLLSESELRPLPHYLRSKALE
SEG  .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
PRD  ecccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  LTTTMMELGALRLQYLCHKYIFYRRLQSLRQEAINHVQIMKETEASYKAQNLYIFLENID
SEG  .....
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ  RLQSLRLQAWTDKQKGLEEKHRECLSSMVTMFPKLQLEWNVHLNIPEVTSPPKPKCKLPA
SEG  .....
PRD  hhhhhhhhhhhhhccccccccccccccccccccccccccccccccccccccccccccccccccce

SEQ  ASPRHIRPSGPTYKQPFLSRHRACVPLQMARQOGKQMEAVWKTEVASSSYAIEKKTASL
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  PRDQLRGHPDIPRLTLDV
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
```

(No Prosite data available for DKFZphtes3_50n23.1)

(No Pfam data available for DKFZphtes3_50n23.1)

DKFZphtes3_6b21

group: testes derived

DKFZphtes3_6b21 encodes a novel 781 amino acid protein without similarity to human KIAA0256 gene product.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to KIAA0256

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: /map="356.3 cR from top of Chr9 linkage group"

Insert length: 3360 bp

Poly A stretch at pos. 3314, polyadenylation signal at pos. 3300

```
1 GGCAAGCCGA CGGCCCGCTG CTGGCCTCCG TGACGCGGCC TCCTCCGCGC
51 CTCGCGGCAT GCGCTCGGAG GGGCCGCGGG AGCCCGAAAG CGAGGGGCATC
101 AAGTTATCAG CAGATGTCAA ACCATTGTG CCCAGATTG CCGGGCTCAA
151 TGTGGCATGG TTAGAGTCCT CAGAAGCATG TGTCTTCCCC AGCTCTGCAG
201 CCACATACTA TCCGTTTGT CAGGAACCA CAGTGACAGA AATGTTTACT
251 CAGTGCCTGG CTCCAGTAT CTTTATAACC AACCCAGTTG TTACCGAGGT
301 TTCAAACAG TGAAGCATCG AAATGAGAAC ACATGCCCTC TCCACAAGA
351 AATGAAAGCT CTGTTTAAGA AGAAAACCTA TGATGAGAAA AAAACGTATG
401 ATCAGCAAAA GTTTGACAGT GAAAGGGCTG ATGGAATAT ATCATCTGAG
451 ATAAATCAG CTAGAGGTTT ACATCATTTG TCCATTTACG CTGAGAATAG
501 TTTGAAATCA GATGGTTACC ATAAGCGAAC AGACAGGAAA TCCAGAATCA
551 TTGCAAAAAA TGTATCTACC TCCAAACCTG AGTTTGAATT TACCACACTG
601 GACTTTCCCTG AACTGCAAGG TGCAGAGAAC AATATGTCAG AGATACAGAA
651 GCAACCCCAAG TGGGGACCTG TCCACTCTGT CTCTACCGAC ATTTCTCTTC
701 TAAGAGAAGT AGTAAAACCA GCTGCAGTGT TATCAAAGGG TGAAATAGTG
751 GTGAAAAATA ACCCAATGA ATCTGTAAT GCTAATGCCG CTACCAATTC
801 TCCTTCATGT ACAAGAGAGT TATCTTGGAC ACCAATGGGT TATGTTGTTT
851 GACAGACATT ATCTACAGAA CTGTCAGCAG CCCCTAAAAA TGTTACTTCT
901 ATGATAAACT TAAAGACCAT TGCTTCATCA GCAGATCCTA AAAATGTTAG
951 TATACCATCT TCTGAAGCTT TATCTTCGGA TCCTTCCTAC AACAAAGAAA
1001 AACACATTAT TCATCCTACC CAAAAGTCTA AAGCATCACA AGGTAGTGAC
1051 CTTGAACAAA ATGAAGCCTC AAGAAAGAAT AAGAAAAAGA AAGAAAAATC
1101 TACATCAAAA TATGAAGTCC TGACAGTTCA AGAGCCTCCA AGGATTGAAG
1151 ATGCCGAGGA ATTTCCCAAC CTGGCAGTTG CATCTGAAAG AAGAGACAGA
1201 ATAGAGACAC CGAAATTTCA ATCTAAGCAG CAGCCACAGG ATAATTTTAA
1251 AAATAATGTA AAGAAGAGCC AGCTTCCAGT GCAGTTGGAC TTGGGGGGCA
1301 TGCTGACAGC CCTGGAGAAG AAGCAGCACT CTCAGCATGC AAAGCAGTCC
1351 TCCAACCCCA AGGTAGTCTC AGTTGGAGCA GTGCCAGTCC TTTCCTTCTC
1401 ATGTGCATCA GGGGAGAGAG GCCGCCGCAT GAGTCAAATG AAGACCCCGC
1451 ACAATCCCTT GGACTCCAGC GCCCCTCTGA TGAAGAAAGG GAAGCAGAGG
1501 GAGATCCCCA AGGCCAAGAA GCCAACCTCA CTGAAGAAGA TTATTTTGAA
1551 AGAACGGCAA GAGAGAAAGC AGCGTCTCCA AGAAAATGCT GTGAGTCCAG
1601 CTTTTACCAG TGATGACACA CAAGATGGAG AGAGTGGTGG TGATGACCAG
1651 TTTCCCGAGC AGGCAGAGCT GTCAGGGCCA GAGGGGATGG ACGAAGTATG
1701 CTCCACTCCT TCGGTTGAGG ACAAGTCTGA AGAGCCACCA GGCACAGAGC
1751 TCCAGAGGGA CACAGAGGCC TCCCACCTTG CTCCCAATCA CACCACCTTC
1801 CCTAAGATCC ACAGCCGAGC ATTCAGGGAT TACTGCAGCC AGATGCTTAG
1851 TAAAGAAGTG GATGCTTGTG TTACCGACCT ACTCAAAGAA CTGGTCCGTT
1901 TCCAAGACCG TATGTACCAG AAAGATCCAG TCAAGGCCAA GACTAAACGT
1951 CGACTTGTGT TGGGGTTGAG GGAGGTCTCT AAACACCTGA AGCTCAAAAA
2001 ACTGAAATGT GTCAATTATT CTCCCAACTG TGAGAAGATA CAGTCAAAAG
2051 GTGGGCTGGA TGACACTTTG CACACAATTA TTGATTATGC CTGTGAGCAG
2101 AACATTCCCT TTGTGTTTGC TCTCAACCGC AAAGCTCTGG GCGCAGTTT
2151 GAATAAGGCA GTTCTGTGTA GTGTGGTGGG GATCTTCAGC TATGATGGGG
2201 CCCAGATCCA GTTCCACAAG ATGGTTGAGC TGACAGTGGC GGCCCGACAG
2251 GCGTACAAGA CCATGCTGGA GAATGTGCAG CAGGAGCTGG TGGGAGAGCC
2301 CAGGCCTCAG GCACCTCCCA GCCTACCCAC ACAGGGCCCC AGCTGCCCTG
2351 CAGAAGATGG CCCCCAGCC CTGAAGAGAA AAGAAGAGCC ACACTACATT
2401 GAAATCTGGA AAAAACATCT GGAAGCATAC AGTGGATGTA CCCTGGAGCT
2451 AGAAGAAATC TTGGAGGCTT CAACCTCTCA AATGATGAAT TTGAATTTAT
2501 GAGAGTTCTT GCCTGTGTGT CTGTATTTTG GGTAAAGGAG GGAGGTCTGA
2551 AAAAGACTTT GGGGCTTTTT CTCTGTTTT TCATGACAA GTAAATTTGTG
2601 TAAGTGTGTA ATCTGGAAT TGATCAGCAT TAAAGGGCAC ATGAAGCAGT
2651 GTCTGCAGCG GTTCAGTGCT GCGGAGCCTG TTAAAGGTCA CTCAGATGTG
```

```

2701 CAGGTGTTAA TCTTCTCTAA AAGCCTGGTT ATACAGCTCT GGCTTTCTGA
2751 GCACACTACG GATCTGGAAA ATACTGGAAA ATGTGATACT TAGAATACTT
2801 TGGCTGCTAA GGAACTTCC TCTCCATTGC AGAATAGCTG AGCCAAGTGA
2851 GTGAGTTTGC AGAAAGCAGG TGGTGAGCTC CTGCCTGCTG GAGGTTGCCA
2901 TGGAGGGCCA TTCCTGCCCC GCAACAGCAC CGTCCTGCAG GGAGCCACTT
2951 GGCAGAAGGG TGCAGGGCTG CTGGTGTCAG AGCAAGAGGG CTACAGGGAA
3001 AGGGCCCTTT CTCAGGGGAT GTAGCTTTTT TAAAAGATTT GGGAACTT
3051 GGAGGATTTG CTAAAATGAG CCTCAGAAGG AAAATTGGTT TTCTAACCTG
3101 TGACTTTTTG AAATGAATTA TTCCTTTCAG TCTTTATTTT TCAAAGAAAC
3151 AATGTGTATT GAAGTACCTA GATTTGTTTG ATAATCAACA AATCTTTCCT
3201 TTTTCAATGA ACATATTCTG AATGTGGTTT CTGTCTTAGA CCAGGAGGAC
3251 AGAGTTTGCT TTCATATTTT CCCTGTAAAGT AAGAGGGCTT ATTTATTTTA
3301 AATAAAGAGT AATTATTAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3351 AAAAAAAAAA

```

BLAST Results

Entry HS773347 from database EMBL:
human STS WI-18160.
Score = 813, P = 2.9e-30, identities = 167/171

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 157 bp to 2499 bp; peptide length: 781
Category: similarity to known protein

```

1 MVRVLRSMCL PQLCSHILSV CSGTTSDRNV YSVPGSQYLY NQPSCYRGFQ
51 TVKHRNENTC PLPQEMKALF KKITYDEKKT YDQKQFDSEK ADGTISSEIK
101 SARGSHHLSI YAENSLKSDG YHKRTDRKSR IIAKNVSTSK PEFEFITLDF
151 PELQGAENNM SEIQKQPKWG PVHSVSTDIS LLREVVKPAA VLSKGEIVVK
201 NNPNESVTAN AATNSPSCTR ELSWTPMGYV VRQTLSTELS AAPKNVTSMI
251 NLKTIASSAD PKNVSIPSE ALSSDPSYNK EKHIHPTQK SKASQGSdle
301 QNEASRNKK KKEKSTSKYE VLTQVEPPRI EDAAEFNLA VASERRDRIE
351 TPKFQSQKQP QDNFKNNVKK SQLPVQLDLG GMLTALEKKQ HSQHAQSSK
401 PVVVSVGAVP VLSKECASGE RGRMSQMKT PHNPLDSSAP LMKKGKQREI
451 PKAKKPTSLK KIILKERQER KQRLQENAVS PAFTSDDTQD GESGGDDQFP
501 EQAELSGPEG MDELISTPSV EDKSEEPGPT ELQRDTEASH LAPNHTTFPK
551 IHSRRFRDYC SQMLSKEVDA CVDLLKELV RFQDRMYQKD PVKAKTKRRL
601 VLGLREVLKH LKLLKLCVI ISPNCEDIQS KGGLLDTLHT IIDYACEQNI
651 PFVFALNRKA LGRSLNKAVP VSVVGIFSYD GAQDQFHKMV ELTVAARQAY
701 KTMLENVQOE LVGEPRPQAP PSLPTQGPSC PAEDGPPALK EKEEPHYIEI
751 WKKHLEAYSG CTLELESLE ASTSQMMNLN L

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_6b21, frame 1

SWISSPROT:Y256 HUMAN HYPOTHETICAL PROTEIN KIAA0256., N = 1, Score = 786, P = 3.6e-78

TREMBL:PFMAL3P3_15 gene: "MAL3P3.15"; Plasmodium falciparum MAL3P3, N = 2, Score = 161, P = 5.1e-10

TREMBL:RNNFLH_1 Rat heavy neurofilament subunit (NF-H) mRNA, 3' end., N = 1, Score = 150, P = 9.1e-07

>SWISSPROT:Y256 HUMAN HYPOTHETICAL PROTEIN KIAA0256.
Length = 635

HSPs:

Score = 786 (117.9 bits), Expect = 3.6e-78, P = 3.6e-78
Identities = 190/424 (44%), Positives = 263/424 (62%)

Query: 369 KKSQLPVQLDLGGLTALAEKKQHSQHAQ--SSKPVVSVGAVPVLSKECASGERGRMS 426
 KK++ PVQLDLG ML ALEK+Q + A+Q +++P+ +V + ++ + S
 Sbjct: 16 KKNKTPVQLDLGDLMLAALEKQQQAMKARQITNTRPLSYTVVTAASFHTKOSTNRKPLTKS 75

Query: 427 Q-MKTPHNPLDSSAPLMKKGKQREIPKAKKPTSLKKIILKERQERKQRLQENAVSPAFTS 485
 Q T N +D ++ KKGK++EI K K+PT+LKK+ILKER+E+K RL + S
 Sbjct: 76 QPCLTSFNSVDIASSKAKGKEKEIAKLKRPTALKKVKILKEREKKGRLTVD--HNLLGS 133

Query: 486 DDTQDGESGGDDQFPEQAELSGPEGMDLISTPSVEDKSEEPG--TELQRDTEASHL-- 541
 ++ + D P++ G+ + S S+ S+ P T + + + AS
 Sbjct: 134 EEPTEMHLDIFIDDLPEIVSQEDTGLS-MPSDTSLSPASQNSPYCMTFVSQGSFASGIG 192

Query: 542 APN-HTTFPKIHSRRFRDYCSQMLSKEVDACVTDLKELVRFQDRMYQKDPVKAKTKRRL 600
 +P +T KIHS+RFR+YC+Q+L KE+D CVT LL+ELV FQ+R+YOKDPV+AK +RRL
 Sbjct: 193 SPMASSTITKIHSKRFRFYCNQVLCKEIDECVTLLQLVLSFQERIYQKDPVRAKARRRL 252

Query: 601 VLGLREVLKHLKLLKLCVLIISPNCEKIQSKGGLDDTLHTIIDYACEQNIPIFVFALNRKA 660
 V+GLREV KH+KL K+KCVIISPNCEKIQSKGGLD+ L+ +I A EQ IPFVFAL RKA
 Sbjct: 253 VMGLREVTKHMKLNKIKCVIISPNCEKIQSKGGLDEALYNVIAMAREQEIPFVFALGRKA 312

Query: 661 LGRSLNKAVPVSVVGIFSYDGAQDFHKMVELTVAARQAYKTMLENVQQLVGEPRP--- 717
 LGR +NK VPVSVVGIF+Y GA+ F+K+VELT AR+AYK M+ ++QE E
 Sbjct: 313 LGRCVNLKVPVSVVGIFNYFGAESLFNKLVELTEEARAYKDMVAAMEQEQAEEALKNVK 372

Query: 718 QAPPSLP-TQGPS-----CPAEDGPPALKEKEPHYIEIWKHLEAYSGCTL---ELE 766
 + P + ++ PS C P + E E Y W+ +E G E E
 Sbjct: 373 KVPHHMHSRNPASAASISFCSEIPEF--ISEVNEKEYETNWRNMVETSDGLEASENEKE 430

Query: 767 ESLEASTSQ 775
 S + STS+
 Sbjct: 431 VSCKHSTSE 439

Pedant information for DKF2phtes3_6b21, frame 1

Report for DKF2phtes3_6b21.1

[LENGTH] 781
 [MW] 87393.44
 [pI] 8.94
 [HOMOL] SWISSPROT:Y256_HUMAN HYPOTHETICAL PROTEIN KIAA0256. 4e-75
 [PROSITE] MYRISTYL 4
 [PROSITE] AMIDATION 1
 [PROSITE] CAMP_PHOSPHO_SITE 3
 [PROSITE] CK2_PHOSPHO_SITE 16
 [PROSITE] TYR_PHOSPHO_SITE 4
 [PROSITE] PKC_PHOSPHO_SITE 16
 [PROSITE] ASN_GLYCOSYLATION 6
 [KW] Alpha Beta
 [KW] LOW_COMPLEXITY 8.45 %

SEQ MVRVLRSMCLPQLCSHILSVCSGTTSDRNVYVPGSQYLYNQPSCYRGFQTVKHRNENTC
 SEG
 PRD cccceeecc

SEQ PLPQEMKALFKKKTDEKKTQYDQKFDSEADGTISSEIKSARGSHHLSIYAENSLKSDG
 SEGxxxxxxxxxxxxx.....
 PRD cccchhhhhhhhhccchhhhhhhhhhhcccccchhhhhhhccceeeeeecccccc

SEQ YHKRTDRKSRIIAKNVSTSKPEFEFTTLDFFELQGAENNMSEIQKPKWGPVHSVSTDIS
 SEG
 PRD cccccchhhheeeccccccccccccccccccccccccchhhhhccccccccceecch

SEQ LLREVVKPAAVLSKGEIVKNNPNESVTANAATNSPCTRELSWTMGYVVRQTLSTELS
 SEG
 PRD hhhhhhheeecc

SEQ AAPKNVTSMLNKTIASSADPKNVISIPSEALSSDPYNKEKHIHPTQKSKASQGSdle
 SEG
 PRD cccccceeehhhhhhccccccccccccccccccccccccccccceehhhhhhhccccch

SEQ QNEASRNKKKKKSTSKYEVLTVQEPRIEDAEFPNLAVASERRDRIETPKFQSKQPP
 SEGxxxxxxxxxxxxx.....
 PRD hhhccccccccccccccccccccccccchhhhhhhccchhhhhhhhhhhcccccccccc

SEQ QDNFKNNVKKSQLPVQLDLGGLTALAEKKQHSQHAQSSKPVVSVGAVPVLSKECASGE
 SEGxxxxxxxxxxxxxxxxxxxxx.....
 PRD cccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhccceeeeeecccccc

```

SEQ  RGRMSQMKTPHNPLDSSAPLMKKGKQREIPKAKKPTSLKKIILKERQERKQRLQENAVS
SEG  .....
PRD  chhhhhccccccccccccchhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhcc

SEQ  PAFTSDDTDQGESGGDDQFPEQAELSGPEGMDLISTPSVEDKSEPPGTQLQDTEASH
SEG  .....
PRD  cccccccccccccccccchhhhhhhccccceeecccccccccccccccccccccccccc

SEQ  LAPNHTTFFPKIHSRRFRDYCSQMLSKEVDACVTDLKELVRFQDRMYQKDPVKAKTKRRL
SEG  .....
PRD  cccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccchhhhhhhhh

SEQ  VLGLREVLKHLKLLKLCVVISPNCEKIQSKGGLDDTLHTIIDYACEQNIPFVFALNRKA
SEG  .....XXXXXXXXXX.....
PRD  hhhhhhhhhhhhhhhheeeccccccccccccchhhhhhhhhhhhhhhccccceeecccccc

SEQ  LGRSLNKAVPVSVVGIFSYDGAQDQFHKMVELTVAARQAYKTMLENVQQELVGEPRPQAP
SEG  .....
PRD  cccccccceeeeeeeccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccc

SEQ  PSLPTQGSPCPAEDGPPALKEKEEPHYIEIWKKHLEAYSGCTLEESLEASTSQMMNLN
SEG  .....XXXXXXXXXXXXXXXXXXXX.....
PRD  cccccccccccccccccchhhhhhhccccceehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccc

SEQ  L
SEG  .
PRD  c

```

Prosites for DKFZphtes3_6b21.1

PS00001	135->139	ASN_GLYCOSYLATION	PDOC00001
PS00001	159->163	ASN_GLYCOSYLATION	PDOC00001
PS00001	204->208	ASN_GLYCOSYLATION	PDOC00001
PS00001	245->249	ASN_GLYCOSYLATION	PDOC00001
PS00001	263->267	ASN_GLYCOSYLATION	PDOC00001
PS00001	544->548	ASN_GLYCOSYLATION	PDOC00001
PS00004	71->75	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	423->427	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	454->458	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	26->29	PKC_PHOSPHO_SITE	PDOC00005
PS00005	51->54	PKC_PHOSPHO_SITE	PDOC00005
PS00005	88->91	PKC_PHOSPHO_SITE	PDOC00005
PS00005	101->104	PKC_PHOSPHO_SITE	PDOC00005
PS00005	115->118	PKC_PHOSPHO_SITE	PDOC00005
PS00005	125->128	PKC_PHOSPHO_SITE	PDOC00005
PS00005	138->141	PKC_PHOSPHO_SITE	PDOC00005
PS00005	288->291	PKC_PHOSPHO_SITE	PDOC00005
PS00005	305->308	PKC_PHOSPHO_SITE	PDOC00005
PS00005	316->319	PKC_PHOSPHO_SITE	PDOC00005
PS00005	343->346	PKC_PHOSPHO_SITE	PDOC00005
PS00005	351->354	PKC_PHOSPHO_SITE	PDOC00005
PS00005	398->401	PKC_PHOSPHO_SITE	PDOC00005
PS00005	458->461	PKC_PHOSPHO_SITE	PDOC00005
PS00005	553->556	PKC_PHOSPHO_SITE	PDOC00005
PS00005	596->599	PKC_PHOSPHO_SITE	PDOC00005
PS00006	24->28	CK2_PHOSPHO_SITE	PDOC00006
PS00006	74->78	CK2_PHOSPHO_SITE	PDOC00006
PS00006	139->143	CK2_PHOSPHO_SITE	PDOC00006
PS00006	146->150	CK2_PHOSPHO_SITE	PDOC00006
PS00006	193->197	CK2_PHOSPHO_SITE	PDOC00006
PS00006	257->261	CK2_PHOSPHO_SITE	PDOC00006
PS00006	297->301	CK2_PHOSPHO_SITE	PDOC00006
PS00006	317->321	CK2_PHOSPHO_SITE	PDOC00006
PS00006	323->327	CK2_PHOSPHO_SITE	PDOC00006
PS00006	384->388	CK2_PHOSPHO_SITE	PDOC00006
PS00006	484->488	CK2_PHOSPHO_SITE	PDOC00006
PS00006	493->497	CK2_PHOSPHO_SITE	PDOC00006
PS00006	506->510	CK2_PHOSPHO_SITE	PDOC00006
PS00006	519->523	CK2_PHOSPHO_SITE	PDOC00006
PS00006	640->644	CK2_PHOSPHO_SITE	PDOC00006
PS00006	702->706	CK2_PHOSPHO_SITE	PDOC00006
PS00007	581->588	TYR_PHOSPHO_SITE	PDOC00007
PS00007	740->748	TYR_PHOSPHO_SITE	PDOC00007
PS00007	740->748	TYR_PHOSPHO_SITE	PDOC00007
PS00007	73->82	TYR_PHOSPHO_SITE	PDOC00007
PS00008	93->99	MYRISTYL	PDOC00008
PS00008	155->161	MYRISTYL	PDOC00008
PS00008	380->386	MYRISTYL	PDOC00008

WO 01/12659

PCT/IB00/01496

PS00008	633->639	MYRISTYL	PDOC00008
PS00009	421->425	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3_6b21.1)

DKFZphtes3_6c11

group: signal transduction

DKFZphtes3_6c11 encodes a novel 1025 amino acid protein with similarity to A. ambisexualis antheridiol steroid receptor.

The novel protein is a putative steroid receptor. It shares similarity with yeast YNL132w and contains the ATP/GTP-binding site motif A (P-loop) and RGD site, similar to the A. ambisexualis antheridiol steroid receptor.

The new protein can find application in modulating/blocking the expression of genes controlled by this receptor.

strong similarity to YNL132w

strong similarity to S.pombe/YDK9_SCHPO, S.cerevisiae/YNL132w, C.elegans/F55A12.8

Sequenced by BMFZ

Locus: unknown

Insert length: 3966 bp

Poly A stretch at pos. 3890, polyadenylation signal at pos. 3873

```
1 GCTGTGCCTT CTCCTTCGGA GTTGTTCCTG GCTCCACAGT GCTTCCCCTT
51 CTCCACTGGC TGGGATCCCC CGGGCTCGGG GCGCAGTAAT AATTTTTCAC
101 CATGCATCGG AAAAAGGTGG ATAACCGAAT CCGGATTCTC ATTGAGAATG
151 GAGTAGCTGA GCGGCAAAGA TCTCTCTTTG TTGTAGTTGG GGATCGAGGA
201 AAAGATCAGG TGGTAATACT TCATCACATG TTATCCAAAG CAACTGTGAA
251 GGCTCGGGCT TCACTGCTGT GGTGTTATAA GAAAGAGCTG GGGTTTAGCA
301 GTCACCGGAA GAAAGAAATG CGACAGCTGC AGAAGAAAAT AAAGAATGGA
351 ACACTGAACA TAAAGCAGGA CGACCCCTTT GAACCTTCA TAGCAGCCAC
401 AAACATTGCG TACTGCTACT ACAACGAGAC CCACAAGATC CTGGGCAATA
451 CCTTCGGGAT GTGTGTGCTG CAGGATTTTG AAGCCTTAAC TCCAAACTTG
501 CTGGCCAGGA CTGTAGAAAC AGTGGAAAGT GGTGGGCTAG TGGTCATCCT
551 CCTACGGACC ATGAACCTAC TCAAGCAATT GTACACAGTG ACTATGGATG
601 TGCATTCCAG GTACAGAAGT GAGGCCATC AGGATGTGGT GGGAAAGATT
651 AATGAAAGGT TTATCTGTCT TCTGGCCTCT TGTAAGAAGT GTCTCGTCAT
701 TGATGACCAG CTCACATCC TGCCCATCTC CTCCACGTT GCCACCATGG
751 AGGCCCTGCC TCCCAGACT CCGGATGAGA GTCTTGGTCC TTCTGATCTG
801 GAGCTGAGGG AGTTGAAGGA GAGCTTGCA GACACCCAGC CTGTGGGTGT
851 GTTGGTGAC TGCTGAAGA CTCTAGACCA GGCCAAAGCT GTCTTGAAAT
901 TTATCGAGGG CATCTCTGAA AAGACCCTGA GGAGTACTGT TGCACTCACA
951 GCTGCTCCAG GACGGGGAAA ATCTGCAGCC CTGGGATTGG CGATTGCTGG
1001 GGGCGTGGCA TTTGGGTACT CCAATATCTT TGTTACCTCC CCAAGCCCTG
1051 ATAACCTCCA TACTCTGTTT GAATTTGTAT TTAAGGATT TGATGCTCTG
1101 CAATATCAGG AACATCTGGA TTATGAGATT ATCCAGTCTC TAAATCCTGA
1151 ATTTAACAAA GCAGTGATCA GAGTGAATGT ATTTGAGAAA CACAGGCAGA
1201 CTATTCAGTA TATACATCCT GCAGATGCTG TGAAGCTGGG CCAGGCTGAA
1251 CTAGTTGTGA TTGATGAAGC TGCCGCCATC CCCCTCCCCT TGGTGAAGAG
1301 CCTACTTGGC CCCTACCTTG TTTTCATGGC ATCCACCATC AATGGCTATG
1351 AGGCACTGG CCGGTCACTG TCCCTCAAGC TAATTCAGCA GCTCCGTCAA
1401 CAGAGCGCCC AGAGCCAGGT CAGCACCCTG GCTGAGAATA AGACCACGAC
1451 GACAGCCAGA TTGGCATCAG CGCGGACACT GCATGAGGTT TCCCTCCAGG
1501 AGTCAATCCG ATACGCCCTT GGGGATGCAG TGGAGAAAGT GCTGAATGAC
1551 TTGCTGTGCC TGGATTGCCT CAACATCACT CGGATAGTCT CAGGCTGCCC
1601 CTTGCCTGAA GCTTGTGAAC TGTACTATGT TAATAGAGAT ACCCTCTTTT
1651 GCTACCACAA GGCTCTGAA GTTTTCTTCC AACGGCTTAT GGCCCTCTAC
1701 GTGCTTCTC ACTACAAGAA CTCTCCCAAT GATCTCCAGA TGCTCTCCGA
1751 TGCACCTGCT CACCATCTCT TCTGCCTTCT GCCTCCTGTG CCCCCACCC
1801 AGAATGCCCT TCCAGAAAGT CTTGCTGTTA TCCAGGTGTG CCTTGAAGGG
1851 GAGATTTCTC GCCAGTCCAT CTTGAACAGT CTGCTCGAG GCAAGAAGGC
1901 TTCAGGGGAC CTGATTCCAT GGACAGTGTC AGAACAGTTC CAAGATCCAG
1951 ACTTTGGTGG TCTGTCTGGT GGAAGGGTCG TTCGATTGCT TGTTCACCCA
2001 GATTATCAAG GGATGGGCTA TGGCAGCCGT GCTCTGCAGC TGCTGCAGAT
2051 GTACTATGAA GGCAGGTTTC CTTGTCTGGA GGAAGAGTTC CTTGAGACAC
2101 CACAGGAAAT TCACACCGTA AGCAGCGAGG CTGTCAGCTT GTTGAAGAG
2151 GTCATCACTC CCCGGAAGGA CCTGCCTCCT TACTCCTCA AATTGAATGA
2201 GAGGCTTCCC GAACGCCTGG ATTACCTGGG TGTTTCTTAT GGCTTGACCC
2251 CCAGGCTCCT CAAGTTCTGG AAACGAGCTG GATTGTGTCC TGTATCTCTG
2301 AGACAGACCC CGAATGACCT GACCGGAGAG CACTCGTGCA TCATGCTGAA
2351 GACGCTCACT GATGAGGATG AGGCTGACCA GGGAGGCTGG CTTGAGCCTT
2401 TCTGGAAGAA TTTCCGACGG CGGTCTCTAG CCTTGCTCTC CTACAGTTC
2451 AGTACCTTCT CTCCTTCCCT GGCTCTGAAC ATCATTCAGA ACAGGAACAT
2501 GGGGAAGCCA GCCCAGCCTG CCCTGAGCCG GGAGGAGCTG GAAGCACTCT
```

```

2551 TCCTCCCCTA TGACCTGAAG CGGCTGGAGA TGTATTCACG GAATATGGTG
2601 GACTATCACG TCATCATGGA CATGATCCCG GCCATCTCTC GCATCTATTI
2651 CCTGAACCAG CTGGGGGACC TGGCCCTGTC TCGGGCTCAG TCGGCTCTTC
2701 TCTTGGGGAT TGGCCTGCAG CATAAGTCTG TGGACCAGCT GGAAAAGGAG
2751 ATTGAGCTGC CCTCGGGCCA GTTGATGGGA CTTTCAACC GGATCATCCG
2801 CAAAATTGTG AAGCTATTTA ATGAAGTTCA GGAAAGGCC ATTGAGGAGC
2851 AGATGGTGGC AGCGAAGGAT GTGGTCATGG AGCCACGAT GAAGACCCTC
2901 AGTGACGACC TAGATGAAGC AGCAAAGGAA TTTCAGGAGA AACACAAGAA
2951 GGAAGTAGGG AAGCTGAAGA GCATGGACCT CTCTGAATAC ATAATCCGTG
3001 GGGACGATGA AGAGTGGAAT GAAGTTTGA ACAAAGCTGG GCCGAACGCC
3051 TCGATCATCA GCCTGAAAAG TGACAAGAAA AGGAAGTTAG AGGCCAAACA
3101 AGAACCCAAA CAGAGCAAGA AGTTGAAGAA CAGAGAGACA AAGAACAAAA
3151 AAGATATGAA ACTGAAGCGG AAGAAATAGT GAAGAGAAAC TCGGGCATCT
3201 GTGTTTGATC ATGGGAAGAT ACTCTCACTA ACTGAACCCT CTCTGGCTGG
3251 ACTGTTAAAA GCAACGAGAG GCCCCGGCAC ACCTGGAAGC TGGCCGCGAA
3301 TTCGGCCTCT GGGCCTGTGT GTCTGTGAGC TCAACCTGGC TAAAGCAGA
3351 GTCACCTCCA AATGGGTCTC TTTAGAACTT GATGGCTGGG CACTGCCATC
3401 TCTAGAATTG CCACGAGTCT CTCTCTTCTT GCCCAGTCCA GGGCCCTCCT
3451 TTCCTATAAG TTCATATTTT GCTTTGAGCC AGCTTTTGTG TCTCATTTCC
3501 ACACATGTGG AAGCCACGTT GCCTCTCGAC CGCCTGAGGC CCTTAAGTAC
3551 ATCGCTTTCT GGTGGTGCCC AGGAGGCTGC TGCTGGGCGG CTGGGTCTCT
3601 CTTTGTGGAG TTGTACCTGG AGCAGGAGGA ACTCCAGTCC GTCCCGGCAT
3651 CCATGGCAGC CCGCGGTTAG GTGCGCCAGG GTTTGCTGAT GTTGTCTTGT
3701 GTGTTTCCAC TCTTGGCTCC AGCAGACCCA CTGTCCCAGA AAAGCCTGAT
3751 CCTGTAGTTT ATGTAGAATG CCACATCTGC GTCCTCAAGA CCTGTTTCAT
3801 CCATTGGGA AAAGATGTTG GGAAAGGCCA CTTTGCTCGC AGGGGTGAGG
3851 GGAAGGATAG AGAATCTATT TTAATAAAT AACATTCTAG AATGAAAAAA
3901 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3951 AAAAAAAAAA AAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 102 bp to 3176 bp; peptide length: 1025
 Category: similarity to unknown protein
 Classification: unclassified
 Prosite motifs: RGD (966-969)
 ATP_GTP_A (284-292)

```

1 MHRKKVDNRI RILIENGVAE RQRSLFVVVG DRGKDQVVIL HHMLSKATVK
51 ARPSVLWCYK KELGFSSHRK KRMRLQKKI KNGTLNIQD DPFEFIAAT
101 NIRYCYNET HKILGNTFGM CVLQDFEALT PNLLARTVET VEGGGLVVIL
151 LRTMNSLKQL YVTMDVHRS YRTEAHQDVV GRFNERFILS LASCKKCLVI
201 DDQLNILPIS SHVATMEALP PQTDES LGP SDLELRELKE SLQDTQPVGV
251 LVDCKTLTDQ AKAVLKFI EG ISEKTLRSTV ALTAARGRGK SAALGLAIAG
301 AVAFGYSNIF VTSPSPDN LH TLFEFVFKGF DALQYQEHLD YEIIQSLNPE
351 FNKAVIRVNV FREHRQTIQY IHPADAVKLG QAEVLVVIDEA AAIPPLVKLS
401 LLGPYLVFMA STINGYEGTG RSLSLKLIQ LQQSAQSQV STTAENKTTT
451 TARLASARTL HEVSLQESIR YAPGDAVEKW LNDLLCLDCL NITRIVSGCP
501 LPEACELYYV NRDTLFCYHK ASEVFLQRLM ALYVASHYKN SPNDLQMLSD
551 APAHHLFCLL PPVPPTQNAL PEVLAVIQVC LEGEISRQSI LNSLSRGKKA
601 SGDLPWTVS EQFQDPDFGG LSGGRVVRIA VHPDYQGMGY GSRALQLLQM
651 YYEGRFPCL EKVLETPQEI HTVSSEAVSL LEEVITPRKD LPPLLLKLNE
701 RPAERLDYLG VSYGLTPRL KFWKRAGFVP VYLRQTPNDL TGEHSCIMLK
751 TLTDDEADQ GGWLAFAWKD FRRRFLALLS YQFSTFSPSL ALNIIQNRNM
801 GKPAQPALS R EELEALFLPY DLKRLEMYSR NMVDYHLIMD MIPAIIRIYF
851 LNQLGDLALS AAQSALLGI GLQHKSVQDL EKEIELPSQG LMGLFNRIIR
901 KVVKLFEVQ EKAIEEQMVA AKDVMMEPTM KTLSDDLDEA AKEFQEKHKK
951 EVGKIKSMDL SEYIIRGDDE EWNEVLNKG PNASIISLKS DKRRKLEAKQ
1001 EPKQSKKLKN RETKNKKDMK LKRRK

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3_6c11, frame 3

TREMBL:CEAF3130_4 gene: "F55A12.8"; *Caenorhabditis elegans* cosmid
F55A12., N = 1, Score = 2782, P = 1.1e-289

PIR:S55151 probable membrane protein YNL132w - yeast (*Saccharomyces cerevisiae*), N = 2, Score = 2549, P = 3.5e-273

SWISSPROT:YXX1_ACHAM HYPOTHETICAL PROTEIN (FRAGMENT)., N = 1, Score = 1013, P = 3.2e-102

SWISSPROT:YDK9_SCHPO HYPOTHETICAL 116.5 KD PROTEIN C20G8.09C IN
CHROMOSOME I., N = 1, Score = 2843, P = 3.8e-296

>SWISSPROT:YDK9_SCHPO HYPOTHETICAL 116.5 KD PROTEIN C20G8.09C IN CHROMOSOME
I.

Length = 1,033

HSPs:

Score = 2843 (426.6 bits), Expect = 3.8e-296, P = 3.8e-296
Identities = 576/1033 (55%), Positives = 750/1033 (72%)

```

Query:      1 MHRKKVDNRIRILIENGVAERQSRSLFVVVGDRGKDQVVLHHMLSKATVKARPSVLWCYK 60
             M +K +D+RI LI+NG +E+QRS FVVVGDR +DQVV LH +LS++ V ARP+VLW YK
Sbjct:      1 MPKKALDSRIPTLIKNGCQEKQRSFFVVVGDRARDQVVLHLLWSQSKVAARPNVLMWYK 60

Query:     61 KEL-GFSSHRKKRMRQLQKKIKNGTLNLIKQDDPFELFIAATNIRYCYYNETHKILGNTFG 119
             K+L GF+SHRKKR +++K+IK G + +DPFELF + TNIRYCY Y E+ KILG T+G
Sbjct:     61 KDLGFTSHRKKRENKIKKEIKRGIRDPNSEDPFELFCSITNIRYCY Y KESEKILGQTYG 120

Query:    120 MCVLQDFEALTPNLLARTVETVEGGGLVVILLRTMNSLKQLYTVTMDVHSRYRTEAHQDV 179
             M VLQDFEALTPNLLART+ETVEGGG+VV+LL +NSLKQLYT++MD+HSRYRTEAH DV
Sbjct:    121 MLVLQDFEALTPNLLARTIETVEGGGIVVLLHKLNSLKQLYTMMDIHSRYRTEAHSDV 180

Query:    180 VGRFNERFILSLASCKKCLVIDDQLNILPISSHVATMEALPPQTPDES LGPSDLELRELK 239
             RFNERFILSL +C+ CLVIDD+LN+LPIS ++ALPP +++ + ++EL+
Sbjct:    181 TARFNERFILSLGNCENCLVIDDELNVLPISGG-KNVKALPPTLEEDN--STQNSIKELQ 237

Query:    240 ESLQDTQPVGVLDCCCKTLDQAKAVLKFIGEISEKTLRSTVALTAARGRGKSAALGLAIA 299
             ESL + P G LV KTLDDQA+AVL F+E I EK+L+ TV+LTA RGRGKSAALGLAIA
Sbjct:    238 ESLGEDHPAGALVGVTKTLDQARAVLTFVESIVEKSLKGTVSLTAGRGRGKSAALGLAIA 297

Query:    300 GAVAFGYSNIFVTSPSPDNLHTLFEFVFKGFDALQYQEHLDYEI IQSLNPEFNKAVIRVN 359
             A+A GYSNIF+TSPSP+NL TLFEF+FKGFDAL Y+EH+DY+IIQS NP ++ A++RVN
Sbjct:    298 AAIAGYSNIFITSPSPENLKTLEFIFKGFDALN YEEHVDYDIIQSTNPAYHNAIVRVN 357

Query:    360 VFREHRQTIQYIHPADAVKLGQAEVLVVIDEAAAIP LPLVKSLLGPYLVFMASINGYEGT 419
             +FR+HRQTIQYI P D+ LGQAEVLVVIDEAAAIP LPLV+ L+GPYLVFMASINGYEGT
Sbjct:    358 IFRDHRQTIQYISPEDSNVLGQAEVLVVIDEAAAIP LPLVRKLIGPYLVFMASINGYEGT 417

Query:    420 GRSLSLKLQQLRQOSAQSQVSTTAENKTTTARLASARTLHEVSLQESIRYAPGDAVEK 479
             GRSLSLKL+QQLR+QS S + NK+ + + + S RTL E+SL E IRYA GD +E
Sbjct:    418 GRSLSLKLQQLREQSRI--YSGSGNNKSDSQSHI-SGRTLKEISLDEPIRYAMGDRIEL 474

Query:    480 WLNDDLCLDCLN-ITRIVS-GCPLPEACELYVNRDTLFCYHKASEVFLQRLMALYVASH 537
             WLN LLCLD + ++R+ + G P P C LY V+RDTLF YH SE FLQR+M+LYVASH
Sbjct:    475 WLNKLLCLDAASYVSRMATQGFPHFSECSLYRVSRDTLFSYHPISEAFLQRMMSLYVASH 534

Query:    538 YKNSPNDLQMSDAPAHHLFCLLPVPPTQNALPEVLAVIQVCLEGEISRQSI LNSLSRG 597
             YKNSPNDLQ++SDAPAH LF LLPPV LP+ + VIQ+ LEG ISR+SI+NSLSRG
Sbjct:    535 YKNSPNDLQMSDAPAHQLFVLLPPVDLKNPKLPDPICVIQLALEGSISRESIMNSLSRG 594

Query:    598 KKASGDLIPWTVSEQFQDPDFGGLSGGRVVRIAVHPDYQGMGYGSRALQLLQMYEGRFP 657
             ++A GDLPW +S+QFQD +F L G R+VRIAV P++ MGYG+RA+QLL Y+EG+F
Sbjct:    595 QRAGGDLIPWLISQFQDENFAALGGARIVRIAVSPEHVKMGYGTAMQLLHEYFEGKFI 654

Query:    658 CLEEKVLETPQEIHVSSEAV---SLLEEVITPR--KDLPPLLLKLNERPAERLDYLGVS 712
             E+ + + E + +L E I R K +PPLLKL+E E L Y+GVS
Sbjct:    655 SASEEFKAVKHSLKRIGDEEIENTALQTEKIHVRDAKTMPPLLLKLSELQPEPLHYGVGS 714

Query:    713 YGLTPRLLKFWKRAGFVVPVLRQTPNDLTGEHSCIMLKTLTDEDEADQGGWLAAFWKDFR 772
             YGLTP L KFWKR G+ P+YLRQT NDLTGEH+C+ML+ L D WL AF ++F
Sbjct:    715 YGLTPSLQKFWKREGYCPLYLRQTANDLTGEHTCVMLRVLEGRDSE---WLGAFQAQNFY 770

Query:    773 RRFLALLSYQFSTFSPSLALNIIQNRNMGKP----AQPALSREELEALFLPYDLKRLEMY 828
             RRFL+LL YQF F+ AL+++ N G + L+ EE+ +F YDLKRLE Y
Sbjct:    771 RRFLSLLGYQFREFAAITALSVLDACNNGTKYVVNSTSKLTNEEINNVSFESYDLKRLESY 830

```

Pedant information for DKFZphtes3_6c11, frame 3

Report for DKFZphtes3_6c11.3

```

SEQ      MHRKKVDNRIRIRILIENGVAERQSRSLFVVVGDRGKDQVVI LHHMMSKATVKARPSVLWCYK
SEG      .
PRD      cccccccchhhhhccccccccceeeeeeeccccceeeehhhhhhhhhhhccccceeehhhh

SEQ      KELGFSSHRKKMRQLQKKIKNGTLN IQDDPFELFIATNIRICYYNETHKILGNTFGM
SEG      .
PRD      hhhccccchhhhhhhhhhhhhhhhhccccccccccccceeeccccceeeccccceeeccccce

SEQ      CVLQDFEALTPNLLARTVETVEGGGLVILLRTMNSLKQLYTVTMDVHSRYRTEAHQDVV
SEG      . . . . . xxxxxxxxxxxxxxxxxxx . . . . .
PRD      eehhhhhccccchhhhhhhhhccccceeeeeecchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ      GRFNERFILSLASCKKCLVIDDQLN I LPISSHVATMEALPPQTPDES LGPSDLELRELKE
SEG      .
PRD      hhhhhhhhhhhccccceeeeeeeccccceccccccccccccccccccccccccchhhhhhhhh

SEQ      SLQDTQPGVGLVDCKCTLDOAKAVLK FIEGISEKTLRSTVALTAARGRKSAALGLAIAG
SEG      . . . . . xxxxxxxxxxx . . . . .
PRD      hhccccccceeeehhhhhhhhhhhhhhhhhhhhhhhhhhhhhheeeccccccchhhhhhhhhhh

SEQ      AVAFGYSNIFVTPSPDNLHTLF EFVFKGFDALQYQEHLDY EIIQSLNPEFNKAVIRNVN
SEG      xxx
PRD      hhhccccceeeccccccccchhhhhhhhhhhhhhhhhhhhhhhheeeccccccccceeeehh

SEQ      FREHRQTIQYIHPADAVKLGQAE LVVIDEAAA I PLPLVKSLLGPYLVMFASTINGYEGTG
SEG      .
PRD      hhhhhhhheeeccccccccccccceeehhhhhhccchhhhhhhccccceeeeeecccccccccc

SEQ      RLSLSLKLIQQLRQOASQSVSTTAENK TTTTARLASARTLHEVSLQESIRYAPGDAVEKW
SEG      . . . . . xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx . . . . .
PRD      cchhhhhhhhhhhhhhhhhhhhhccccccccccchhhhhhhhhhhhhhhheeeccccchhhh

SEQ      LNDLLCLDCLNITRIVSGCPLPEACE LYYVNRDTLFCYHKASEVFLQRLMALYVASHYKN
SEG      xxxxxxxxxxx .
PRD      hhhhhccccceeeccccccccceeeeeeccccccccchhhhhhhhhhhhhhhhhhhhhcc

SEQ      SPNDLQMLSDAPAHHLFCLLPVPPTQ NALPEVLAVIQVCEGEISRQSILNSLSRGKKA
SEG      . . . . .
PRD      cccccccccccccceeeeeeccccccccccchhhhhhhhhccccchhhhhhhhhcccccc

SEQ      SGDILPWTVSEQFQDPDFGGLSGGRV VR IAVHPDYQGMGYGSRALQLLQMYYEGRFPCLE
SEG      .
PRD      cccchhhhhhhhhhhccccccccceeeeeeccccccccccchhhhhhhhhhhhhccccchhh

SEQ      EKVLETPOEIHTVSSEAVSLLEEVI TP RKDLPPLLKLNERPAERLDYLGVSYGLTPRL

```

```

SEG      .....XXXXXXXXX.....
PRD      hhhhhccccccchhhhhhhhhhhhhccccccccccccccccccccccccccccccccccccchhh
SEQ      KFWKRAGFVPVYLQTFPNDLTGEHSCIMLKTLTDEDEADQGGWLAAFWKDFRRRFLALLS
SEG      .....
PRD      hhhhhccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhh

SEG      YQFSTFSPSLALNIIQNRNMGKPAQPALSREELEALFLPYDLKRLEMYSRNMVDYHLIMD
SEG      .....
PRD      hhhhccchhhhhhhhhhhccccccccchhhhhhhhhhhhhccchhhhhhhhhccchhhhhhhhh

SEG      MIPAISRIYFLNLGLDALSAAQSALLLGIGLQHKSVQLEKEIELPSGQLMGLFNRIIR
SEG      .....XXXXXXXXXXXXXXXXXXXXX.....
PRD      hhhhhhhhhhhhhccccchhhhhhhhhhhhhccchhhhhhhhhhhhhccccchhhhhhhhhhh

SEG      KVVKLFNEVQEKAIEEQMVAADVVMEPTMKTLSDDLDEAAKEFQEKHKKEVGKCLKSMDL
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhcc

SEG      SEYIIRGDDEEWNEVLNKGPNASIISLKSDDKKRLEAKQEPKQSKKLKNRETKNKKDMK
SEG      .....XXXXXXXXXXXXXXXXXXXXX.....
PRD      cccccccccchhhhhhhhhhhccccccccccccccccchhhhhhhhhhhccccccccccccccccchhh

SEG      LKRKK
SEG      xxxxx
PRD      hhccc

```

Prosite for DKFZphtes3_6c11.3

```

PS00016      966->969      RGD      PDOC00016
PS00017      284->292      ATP_GTP_A      PDOC00017

```

(No Pfam data available for DKFZphtes3_6c11.3)

DKFZphtes3_6d16

group: testes derived

DKFZphtes3_6d16 encodes a novel 695 amino acid protein nearly identical to a sequence from human PAC clone WUGSC:H_DJ1185I07.2.

The cDNA is different to the proposed gene model: it contains additional exons.
No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

WUGSC:H_DJ1185I07.2, differences to genmodel

differences to genmodel of WUGSC:H_DJ1185I07.2 two exons skippt,

Sequenced by BMFZ

Locus: /map="7q11.23-q21"

Insert length: 4572 bp

Poly A stretch at pos. 4540, polyadenylation signal at pos. 4520

```

1  GCGCGCGCTA GCTTCGGAGT CTCCCGCGCG CACCTCAGCC GCCTCCTAGC
51  GCGCGCGCGC TCGCTCCTAC GCCTAAAATG ACCAATGTGT GATTTCAGTG
101 GAATAAATGG CGTCCAAAGT CACAGATGCT ATAGTCTGGT ATCAAAAGAA
151 GATTGGAGCA TATGATCAAC AAATATGGGA AAAATCTGTT GAACAGAGAG
201 AAATCAAGGG GCTAAGGAAT AAACCAAAGA AAACAGCACA TGTGAAACCA
251 GACCTCATAG ATGTTGATCT TGTAAGAGGG TCTGCATTGG CAAAGGCCAAA
301 GCCTGAAAGT CCTTGGACTT CTCTGACCAG AAAGGGAATT GTTCGAGTTG
351 TATTTTTCCT CTTTTCTTTC CGGTGGTGGT TACAAGTAAC ATCAAAGGTC
401 ATCTTTTCTT GGCCTTCTGT CCTTTATCTT CTTCAGTTG CTGCAATAGT
451 ATTATTCTGC TCCACTTCTA GCCCACACAG CATACTCTG ACAGAGGTGA
501 TTGGGCGGAT ATGGCTGATG CTGCTCCTGG GAACTGTGCA TTGCCAGATT
551 GTTTCACCAA GAACACCCAA ACCTCCTCTA AGTACAGGGG GTAAAAGAAG
601 AAGGAAATTA AGAAAAGCAG CCCATTGGGA AGTACATAGG GAAGGAGATG
651 GTTCTAGTAC CACAGATAAC ACACAAGAGG GAGCAGTTCA GAACCACGGT
701 ACAAGCACCT CTCACAGCGT TGGCACTGTC TTCAGAGATC TCTGGCATGC
751 TGCTTTCTTT TTATCAGGAT CAAAGAAAGC AAAGAATTCA ATTGATAAAT
801 CAACTGAAAC TGACAATGGC TATGTATCCC TTGATGGGAA GAAGACTGTT
851 AAAAGCGGTG AAGATGGAAT ACAAACCAT GAACCTCAGT GTGAAACTAT
901 TCGACCAAGG GAGACAGCCT GGAACACAGG AACACTGAGG AATGGTCTCA
951 GCAAAGATAC CCAAAGGACA ATAACAAATG TCTCTGATGA AGTCTCCAGT
1001 GAGGAAGGCT CTGAAACAGG ATACTCATTA CGTCGTCATG TGGACAGGAC
1051 TTCTGAAGGT GTTCTTCGGA ATAGAAAGTC ACACCATTAT AAGAAACATT
1101 ACCCTAATGA GGACGCCCTT AAATCGGGTA CTAGTTGCAG CTCTCGCTGT
1151 TCAAGTTCCA GACAGGATTC TGAGAGTGCA AGGCCAGAAT CTGAAACAGA
1201 AGATGTGTTA TGGGAAGACT TGTTACATTG TGCAGAAATG CATTCATCTT
1251 GTACCAAGTA GACAGATGTG GAAATCATC AGATTAATCC ATGTGTGAAA
1301 AAGAAATATA GAGATGACCC TTTTCATCAG AGTCATTTCG CCTGGCTCCA
1351 TAGTTCACAC CCAGGATTAG AAAAAATAAG TGCTATAGTA TGGGAAGGTA
1401 ATGATTGTAA GAAAGCAGAC ATGTCTGTAC TTGAAATCAG TGGAAATGTA
1451 ATGAACAGAG TGAACAGCCA TATACCAGGA ATAGGATACC AGATTTTGGG
1501 AAATGCAGTC TCTCTCATAC TGGGTTTAAC TCCATTGTGT TTCCGACTTT
1551 CTCAAGCTAC AGACTTGGAA CAACTCACAG CACATTCTGC TTCAGAACTT
1601 TATGTGATTG CATTGGTTTC TAATGAAGAT GTCATAGTTC TTTCTATGGT
1651 TATAATAAGT TTTGTGGTTC GCGTGTCTCT TGTGTGGATT TTCTTTTTTT
1701 TGCTCTGTGT AGCAGAAAGA ACTTATAAAC AGCGATTACT TTTTGCAAAA
1751 CTCTTTGGAC ATTTAACATC TGCAAGGAGG GCTCGAAAAT CTGAGGTTCC
1801 TCATTTCCGG TTGAAGAAAG TACAGAATAT AAAAATGTGG CTATCTCTCC
1851 TTCTCTATCT TAAGCGTCGA GGTCTCAGC GATCAGTTGA TGTAAATAGT
1901 TCACTGCTT TCTTATTGAC TATCTCAGTT GTATTTATCT GTTGTGCCCA
1951 GATAAACCTC TACTTGAAAA TGGAGAAAAA ACCTAACAAA AAGGAGGAAC
2001 TGACACTAGT GAATAATGTT TTAATACTGG CTACTAAACT GCTAAAGGAG
2051 TTGGACAGTC CTTTATAGATT ATATGGGCTT ACAATGAATC CGTGTCTTTA
2101 TAACATCACC CAGGTTGTTA TCCTGTACAG TGTTCCTGGT GTTATCAGTG
2151 ACTTGCCTTG ATTTAATTTA AAGCTATGGA AGATTAAGTC ATGACAATTC
2201 AAAGAAAAGA AGATGTAGCC TCTTTTCCAG AATAAGAGTA CTGACTAAGC
2251 TGCTGAAAG CTGTGCTACT ATTCTTTGCT TCAGGAGTCT CAGCTAGGGA
2301 GTTGAAGTGT TTACATCAGA CTGTCTTGTG CAATTCCTAT ATTTATTTTA
2351 CTGGTTCAC TTTTTTTACA TTTATTTTAG TCTTATATT TTTATTTTAA
2401 AGCATTGATG TACTTAGTTG TTGAAAGGGT GATGAACTG ATATCCAGAT
2451 ACTTGAGATC CTGGTAATTG GTCATAAATA ATTGCAAAA TAACAAATTG
2501 TGAAAATAGA AGCCATTGCT CAGCACCGTT TCTCCATCAA TGCCGTGAAC
2551 TTGCCTTACT TGAGGAAAAA TTCCTTAACT TTGGAATATT GCATTGAACT
2601 CAGCTATACA CATAAAACAT TTTCTTTGGT AAATCAAGAT CCAGTCAGGG

```

```
2651 TTTCTCTTGA ATTATTTTGG AACAAATGCCA GGATCCAAAC TGATTAAGTT
2701 ACAGTTTAAG CACCCTTCAG TATTAATATA TACGGTATTA TATAACAGGT
2751 CAACAAGTGC TCTTTGATGA TAAAACCTGT AATAGAGCAA TAATTGTAAA
2801 TGGTTACCAT ACTGTAAGAT ATTTTGATAA AAATTAACATA GTAATACTTG
2851 TATTTATTTG AAACACTGGG CTGTTTGCAC AGCTCCAACG GTGCATGCTC
2901 AAAATGTGCA CTTTTTAAAA TTGTTACTTT TAATGCGTAT CTTTATATGG
2951 GATCTGTTAT AGTATACTAG GGCATGATAT GGTATCCTTT TGAGTGAGGT
3001 ATATACTCAT CTCACAAGTG AAGTGCCTAC TGATATTACT AAAGTACATT
3051 ATGTTTACTC AAGTAAATAA TTTTCTCCCC ATGGTACACT CTAGTGTAGG
3101 CTATTCATAC CACACTGAAA TGAACAAC TGAAATAAGG CTAAGAACCA
3151 ATAAATATTT TCTCTAATTG CTAGTTGTAA AACTGTATCC AAATTTTCAG
3201 AAAAGACAGC TTCAGCTTGC AAATTTCTATC CTCTAAACTT ATCTGGTGCA
3251 TTCTCCCCAC CCCACCCCCA TTATATAAGG GCTATTTTAG ATGCTTTTAA
3301 CCTCCCCAAC AAATAATTG CCAAGTGTC AATGAGAACT TATCATGTTG
3351 GTGTGTTAGG TAAATCGGGC AAATATGATA GTGTCTTACA TTGGGCCCTG
3401 ATTTTAAGTT GTTATATTG TACAATCGAG TATTTTAGAA ATTACATGAA
3451 ACATGAAACA GTTTTGCAG TTTTTTTAA ACTGGGCATC TGGTTTCTAA
3501 AAATTTATTT GAAACAATCT AGAATTTTCT TGGTGCAAAG TGTATCATGT
3551 GGAATATCCT CATATTTTTA CCATATTTTA AGAACTTTAA GACGATTAAT
3601 TGTAATAAT TTATTGATT GGTGCAGTTC TAATCCCTAA ATCATAATCT
3651 TAAATCAGG AATGTGTGGA GAACAGAGCC ATGTCATATC ACTTTGCTCT
3701 TACCATTCTT TTGATCAGC CTCAATTCAG CCTCATTTGT TAGTATGTTT
3751 TTTCTTTCTA TGAAAAACAA CAGAAAGCAT TTCATTTTAT TTGCCTATGT
3801 TCAATATGT TTAATAATGA CCAAAGTGCA TTCTGAGTTT TTTCAAGGAA
3851 TGTAATACTG GAGCTTTAAG AACATACTTA GTTTCTCATG TGAAAACTTA
3901 GGCTTTGTCT GATGTTTTTC CTTCCTCTAT TGTCTAATGT TGAGGTGTTT
3951 TTTAGGAATT ATGTTTATA AACTTTTCA ATATAAGGTA CATGCCTATA
4001 CAGAACTTAA CATTTTGCAC AGAATATATC AAATATATTT TGAGAAAAAA
4051 AGTACGGCAT GAGTTCTGTT AGGAATAAAA GATGAACTA TTGTATCTCA
4101 CAAAAAATCT TATTTAGAA TGGAAATATT TTTAGAAAAA GTAGCTGAGT
4151 ATACTGGTTT AAGAAATGC TTGTTTAGA TTGAGGTAA CTTAGAGTTG
4201 GGAGTTGATT TATTAAGTAC AGTATACCTC TCAACAGTTT ATAAATAATA
4251 TGTGAATTA TGTCAAGTGT GGCAGCAGTA GAATACTAAA AGGAAATGT
4301 CATGTTAAGC AATTTAGAA CATTAACTGA ACTATTTTCA AAGCAGAAAA
4351 ATTGACATTG CTGCCCTTAA GAATACCATG AATGTAAGAA ATTGAAAGAA
4401 ATTTGAAAT ATCATAAT ATAGAAATGG CAGTTCAAAG AGAATGTGG
4451 CAGATGTTGT GTGTGAAC TGTTTCTTT GCCACATGTG TTGTATTTGA
4501 AAGTTTACA GTAAGTTTAA AATAAAACAT TCTGTGACTG AAAAAA
4551 AAAAAA AAAAAA AA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 107 bp to 2191 bp; peptide length: 695
Category: known protein
Classification: unclassified
Prosite motifs: CYTOCHROME_C (375-381)

```
1 MASKVTDIAV WYQKKIGAYD QQIWEKSVEQ REIKGLRNKP KKTAKHVKPDL
51 IDVDLVRGSA FAKAKPESPW TSLTRKGIVR VVFFPFFFRW WLQVTSKVIF
101 FWLLVLYLQ VAAIVLFCST SSPHSIPLTE VIGPIWMLL LGTVHCQIVS
151 TRTPKPPPLT GKKRRRLRK AAHLEVRHRE DGSSTTDNTQ EGAVQNHGTS
201 TSHSVGTVFR DLWHAAFFLS GSKKAKNSID KSTETDNGYV SLDGKKTIVS
251 GEDGTQNHPE QCETIRPEET AWNTGTLRNG PSKDTQRTIT NVSDEVSSSE
301 GPETGYSLRH HVDRTSEGV LNRKSHHYKK HYPNEDAPKS GTCSSSRCS
351 SRQDESARPE ESETEDVLWE DLLHCAECHS SCTSETDVEN HQINPCVKKE
401 YRDDPFHQSH LPWLHSSHPG LEKISAIWVE GNDCKKADMS VLEISGMIMN
451 RVNSHIPGIG YQIFGNAVSL ILGLTPFVFR LSQATDLEQL TAHSASELYV
501 IAFGSNEDVI VLSMVIISFV VRVSLWIFF FLLCVAERTY KQRLLFKALF
551 GHLSARRAR KSEVPFRLK KVQNIKMWLS LRSYLKRRGP QRSVDVIVSS
601 AFLLTISVVF ICCAQINLYL KMEKKPNKKE ELTLVNNVLK LATKLLKELD
651 SPFRLYGLTM NPLLYNITQV VILSAVSGVI SDLLGFNLKL WKIKS
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_6d16, frame 2

PIR:S38170 SRP40 protein - yeast (*Saccharomyces cerevisiae*), N = 1,
Score = 100, P = 0.08

TREMBL:AC004990_1 gene: "WUGSC:H_DJ1185I07.2"; Homo sapiens PAC clone
DJ1185I07 from 7q11.23-q21, complete sequence., N = 2, Score = 2693, P
= 0

>TREMBL:AC004990_1 gene: "WUGSC:H_DJ1185I07.2"; Homo sapiens PAC clone
DJ1185I07 from 7q11.23-q21, complete sequence.
Length = 588

HSPs:

Score = 2693 (404.1 bits), Expect = 0.0e+00, Sum P(2) = 0.0e+00
Identities = 510/515 (99%), Positives = 512/515 (99%)

```
Query:   35 GLRNKPKKTAHVKPDLDVLDVLRGSFAFAKAKPESPWTSLTRKGIVRVVFFPFFFRWWLQV  94
          GLRNKPKKTAHVKPDLDVLDVLRGSFAFAKAKPESPWTSLTRKGIVRVVFFPFFFRWWLQV
Sbjct:   1  GLRNKPKKTAHVKPDLDVLDVLRGSFAFAKAKPESPWTSLTRKGIVRVVFFPFFFRWWLQV  60

Query:   95 TSKVIFFWLLVLYLLQVAAIVLFCSTSSPHSIPLTEVIGPIWMLLLGTVHCQIVSTRTP  154
          TSKVIFFWLLVLYLLQVAAIVLFCSTSSPHSIPLTEVIGPIWMLLLGTVHCQIVSTRTP
Sbjct:   61 TSKVIFFWLLVLYLLQVAAIVLFCSTSSPHSIPLTEVIGPIWMLLLGTVHCQIVSTRTP  120

Query:   155 KPPLSTGGKRRRKLKRAAHLEVHREGDGSSTTDNTQEGAVQNHGTSTSHSVGTVFRDLWH  214
          KPPLSTGGKRRRKLKRAAHLEVHREGDGSSTTDNTQEGAVQNHGTSTSHSVGTVFRDLWH
Sbjct:   121 KPPLSTGGKRRRKLKRAAHLEVHREGDGSSTTDNTQEGAVQNHGTSTSHSVGTVFRDLWH  180

Query:   215 AAFFLSGSKKAKNSIDKSTETDNGYVSLDGKKTVKSGEDGIQNHQPQCETIRPEETAWNT  274
          AAFFLSGSKKAKNSIDKSTETDNGYVSLDGKKTVKSGEDGIQNHQPQCETIRPEETAWNT
Sbjct:   181 AAFFLSGSKKAKNSIDKSTETDNGYVSLDGKKTVKSGEDGIQNHQPQCETIRPEETAWNT  240

Query:   275 GTLRNGPSKDTQRTITNVSDVSSSEEGPETGYSLRRHVDRTSEGLNRNRKSHHYKKHYPN  334
          GTLRNGPSKDTQRTITNVSDVSSSEEGPETGYSLRRHVDRTSEGLNRNRKSHHYKKHYPN
Sbjct:   241 GTLRNGPSKDTQRTITNVSDVSSSEEGPETGYSLRRHVDRTSEGLNRNRKSHHYKKHYPN  300

Query:   335 EDAPKSGTSCSSRCSSSRQDSSESARPESETEDVLWEDLLHCAECHSSCTSETDVENHQIN  394
          EDAPKSGTSCSSRCSSSRQDSSESARPESETEDVLWEDLLHCAECHSSCTSETDVENHQIN
Sbjct:   301 EDAPKSGTSCSSRCSSSRQDSSESARPESETEDVLWEDLLHCAECHSSCTSETDVENHQIN  360

Query:   395 PCVKKEYRDDPFHQSHLPWLHSSHPGLEKISAIWVEGNDCKKADMSVLEISGMIMNRVNS  454
          PCVKKEYRDDPFHQSHLPWLHSSHPGLEKISAIWVEGNDCKKADMSVLEISGMIMNRVNS
Sbjct:   361 PCVKKEYRDDPFHQSHLPWLHSSHPGLEKISAIWVEGNDCKKADMSVLEISGMIMNRVNS  420

Query:   455 HIPGIGYQIFGNAVSLILGLTPFVFRLSQATDLEQLTAHSASELYVIAFGSNEDVIVLSM  514
          HIPGIGYQIFGNAVSLILGLTPFVFRLSQATDLEQLTAHSASELYVIAFGSNEDVIVLSM
Sbjct:   421 HIPGIGYQIFGNAVSLILGLTPFVFRLSQATDLEQLTAHSASELYVIAFGSNEDVIVLSM  480

Query:   515 VIISFVVRVSLVWIFFFLLCVAERTYKQRLLFACL  549
          VIISFVVRVSLVWIFFFLLCVAERTYKQ L+ K+
Sbjct:   481 VIISFVVRVSLVWIFFFLLCVAERTYKQINLYLKM  515
```

Score = 409 (61.4 bits), Expect = 0.0e+00, Sum P(2) = 0.0e+00
Identities = 92/115 (80%), Positives = 98/115 (85%)

```
Query:   595 DVIVSS----AFLLTISVVFI-----CCA-----QINLYLKMEKKPNKKEELTLVNNVLK  640
          DVIV S  +F++ +S+V+I  C A  QINLYLKMEKKPNKKEELTLVNNVLK
Sbjct:   474 DVIVLSMVIISFVVRVSLVWIFFFLLCVAERTYKQINLYLKMEKKPNKKEELTLVNNVLK  533

Query:   641 LATKLLKELDSPFRLYGLTMNPLLYNITQVVILSAVSGVISDLLGFNLKLWKIKS  695
          LATKLLKELDSPFRLYGLTMNPLLYNITQVVILSAVSGVISDLLGFNLKLWKIKS
Sbjct:   534 LATKLLKELDSPFRLYGLTMNPLLYNITQVVILSAVSGVISDLLGFNLKLWKIKS  588
```

Pedant information for DKFZphtes3_6d16, frame 2

Report for DKFZphtes3_6d16.2

[LENGTH] 695
[MW] 78466.68
[pI] 9.30
[HOMOL] TREMBL:AC004990_1 gene: "WUGSC:H_DJ1185I07.2"; Homo sapiens PAC clone DJ1185I07
from 7q11.23-q21, complete sequence. 0.0

[PROSITE] CYTOCHROME_C 1
[KW] TRANSMEMBRANE 6
[KW] LOW_COMPLEXITY 5.32 %

```
SEQ MASKVTDAlVWYQKKIGAYDQQIWEKSVEQREIKGLRNKPKKTAHVKPDLDVLDVRGSA
SEG .....
PRD cccceeeehhhhhhhccccchhhhhhhhhhhhhhhccccccccccccccccccccccccccch
MEM .....

SEQ FAKAKPESPWTSLTRKGIVRVVFFPFFFRWWLQVTSKVIFFWLLVLYLLQVAAIVLCST
SEG .....XXXXXXXXXX.....
PRD hhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ SSPHSIPLTEVIGPIWMLLLGTVHCQIVSTRTPKPPLSTGGKRRRKLKAAHLEVHREG
SEG .....XXXXXXXXXX.....
PRD cccccccccceehhhhhhhhhhhhhheeeeeeccccccccccccchhhhhhhhhhhhhheeecc
MEM .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ DGSSTTDNTQEGAVQNHGTSTSHSVGTVFRDLWHAFFLSGSKKAKNSIDKSTETDNGYV
SEG .....
PRD cccccccccceeeeeeccccccccchhhhhhhhhhhhhhhccccchhhhhcccccccccccccc
MEM .....

SEQ SLDGKTKVKSGEDGIQNHQPQCETIRPEETAWNTGTLRNGPSKDTORTITNVSEVSSEE
SEG .....
PRD cccccceeecccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM .....

SEQ GPETGYSLRRHVDRTSEGVLNRKSHHYKKHYPNEDAPKSGTSCSSRCSSSRQDESARP
SEG .....XXXXXXXXXXXXXXXXXXXXX...
PRD cccccceeeccccccccchhhhhcccccccccccccccccccccccccccccccccccccccccc
MEM .....

SEQ ESETEDVLWEDLLHCAECHSSCTSETDVENHQINPCVKKEYRDDPFHQSHLPWLHSSHFG
SEG .....
PRD ccccchhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccc
MEM .....

SEQ LEKISAIWEGNDCKKADMSVLEISGMIMNRVNSHIPGIGYQIFGNAVSLILGLTPFVFR
SEG .....
PRD cccccceeeccccccccceeeehhhhhhhhhccccccccccccccccccccccccccccchh
MEM .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ LSQATDLEQLTAHSASELYVIAFGSNEDVIVLSMVIISFVVRVSLVWIFFLLCVAERTY
SEG .....
PRD hhhhhhhhhhhhhccccceeeeeeccccceeehhhhhhhhhhccccchhhhhhhhhhhhhhhhh
MEM .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ KQRLFLAKLFGHLTSARRARKSEVPFRLKKVQNIKMWLSLRSYLKRRGPQRSVDIVVSS
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccc
MEM .....MMMMMM

SEQ AFLLTISVVFICCAQINLYLKMEKKPNKKEELTLVNNVLKLATKLLKELDSPFRLYGLTM
SEG .....
PRD eeeeeeeeeeeehhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhcccccccccccccc
MEM MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ NPLLYNITQVILSAVSGVISDLLGFNLKLWKIKS
SEG .....
PRD cchhhhhheeeeeeccccchhhhhcccccccccccccc
MEM MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....
```

Prosite for DKFZphtes3_6d16.2

PS00190 375->381 CYTOCHROME_C PDOC00169

(No Pfam data available for DKFZphtes3_6d16.2)

DKFZphtes3_72k11

group: testes derived
 DKFZphtes3_72k11 encodes a novel 233 amino acid protein with similarity to S.pombe
 hypothetical repeat-containing protein.

The novel protein contains 5 leucine zippers and a microbodies C-terminal targeting signal (S-K-L) signature. This sequence is responsible for transport of proteins from free polysomes into the microbodies.
 No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to S.pombe hypothetical repeat-containing protein

complete cDNA, complete cds, 6 EST hits (3 from testis derived libraries)

Sequenced by DKFZ

Locus: unknown

Insert length: 1134 bp

Poly A stretch at pos. 1124, polyadenylation signal at pos. 1088

```

1 AACCTTTCAA GTGCCCCCTC CTTTCCTTAA AGTCTTTTAT AGGGGTCCCC
51 TTCTTGGCCA TCTCCATCCT GTGAGTCAGG ACTGAAAGGG CACAGACAGG
101 TCACTGCCAG CATTGTTGGG GCAAGCCTGC AAGCACGCAT CACTGGGGAT
151 CTGACATGAC AATGGCCGCC TGCCCCCTCT GAGGGCTACA GGAATTACCC
201 CAGTGGGAAG CAGCTAAGCA GGTCTGACCA GCCGACCTGG ACCTGGCCAA
251 GGGTCCTGTC ATCCCTCATG GCCACCCCGC CATTCGGGCT GATAAGGAAG
301 ATGTTTCTCT TCAAGGTGAG CAGATGGATG GGGCTTGCCT GCTTCCGGTC
351 CCTGGCGGCA TCCTCTCCCA GTATTGCGCA GAAGAACTA ATGCACAAGC
401 TGCAGGAGGA AAAGGCTTTT CGCGAAGAGA TGAAAATTTT TCGTGAAAAA
451 ATAGAGGACT TCAGGGAAGA GATGTGGACT TTCCGAGGCA AGATCCATGC
501 TTTCGGGGGC CAGATCCTGG GTTTTGGGA AGAGGAGAGA CCTTCTGGG
551 AAGAGGAGAA AACCTTCTGG AAAGAGGAAA AATCCTTCTG GGAAATGGAA
601 AAGTCTTTCA GGGAGGAAGA GAAAACTTC TGGAAAAAGT ACCGCACTTT
651 CTGGAAGGAG GATAAGGCCT TCTGGAAGA GGACAATGCC TTATGGGAAA
701 GAGACCGGAA CCTTCTTCAG GAGGACAAGG CCCTGTGGGA GGAAGAAAAG
751 GCCCTGTGGG TAGAGGAAAG AGCCCTCCTT GAGGGGGAGA AAGCCCTGTG
801 GGAAGATAAA ACGTCCCTCT GGGAGGAAGA GAATGCCCTC TGGGAGGAAG
851 AGAGGGCCTT CTGGATGGAG AACAATGGCC ACGTTGCCGG AGAGCAGATG
901 CTCGAAGATG GGCCCCACAA CGCCAACAGA GGGCAGCGCT TGCTGGCCTT
951 CTCCCGAGGC AGGGCGTAGC CAGCATGCAG GTGCAGGGCC CTGTGGTCCA
1001 GACTCCCCCT GGTGGGATT CAAGTCCAGG GTGAGCCCAT GTGCTGGAGA
1051 AAATACACAC TCATTGGTCT CCTTGCTTTG AAAGATCCAA TAAAGTCCTG
1101 AGGCAAGGTT TGGAAAACCA ACTTAAAAAA AAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 268 bp to 966 bp; peptide length: 233
 Category: similarity to known protein
 Prosite motifs: MICROBODIES_CTER (231-234)
 LEUCINE_ZIPPER (142-164)
 LEUCINE_ZIPPER (149-171)
 LEUCINE_ZIPPER (156-178)
 LEUCINE_ZIPPER (163-185)
 LEUCINE_ZIPPER (170-192)
 LEUCINE_ZIPPER (170-192)

BLASTP hits

Entry A45973 from database PIR:
trichohyalin - human
Score = 147, P = 3.0e-07, identities = 57/194, positives = 94/194

No Alert BLASTP hits found

Pedant information for DKFZphtes3 72k11, frame 1

Report for DKFZphtes3 72k11.1

Prosites for DKFZphtes3 72k11.1

(No Pfam data available for DKFZphtes3 72k11.1)

DKFZphtes3_72k15

group: cell structure and motility

DKFZphtes3_72k15 encodes a novel 188 amino acid protein with strong similarity to *Rattus norvegicus* actin-filament binding protein Frabin.

FGD1-related F-actin-binding protein (Farbin/FGD1) is a novel F-actin-binding protein. The gene locus *fgd1* seems to be responsible for faciogenital dysplasia or Aarskog-Scott syndrome. Frabin binds F-actin and shows F-actin-cross-linking activity. Overexpression of frabin in Swiss 3T3 cells and COS7 cells induces cell shape change and c-Jun N-terminal kinase activation, as described for FGD1. Because FGD1 has been shown to serve as a GDP/GTP exchange protein for Cdc42 small G protein, it is likely that frabin is a direct linker between Cdc42 and the actin cytoskeleton. Cdc42p is an *esin* yeast, Cdc42p transduces signals to the actin cytoskeleton to initiate and maintain polarized growth and to mitogen-activated protein morphogenesis. In mammalian cells, Cdc42p regulates a variety of actin-dependent events and induces the JNK/SAPK protein kinase cascade, which leads to the activation of transcription factors within the nucleus. The novel protein seems to be the human orthologue of rat frabin.

The new protein can find application in modulating of cell structure and motility as well as modulation of the JNK/SAPK pathway.

strong similarity to actin-filament binding protein Frabin

2 EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 1845 bp

Poly A stretch at pos. 1835, polyadenylation signal at pos. 1816

```
1 GTGATGGAGA GTGCTGTTAT GATAGATGAA TCTAGGAAAG CCTCTTTGGA
51 GATGTGATAC CTGAACAGAA CCCCGAATGA TAAGAAGAAA TACCAGTGTT
101 TTAGGAGAGA TTGTCTTAAG CAGAGAACAG CAGCTGCAAA GACCCCAAGA
151 CACATACACT TGGTTATTAA GAATGGGAGC AGCAAGGAGT ATGGCAAGAA
201 .CACAGTGAGT TTTCCCTTGA GTGTGTGAGG AAGCCCTCAG AGTTTGTGAC
251 TGACTTTGTAG AGGTTCTAGT GGAGGGGATC AGAGTGGAAG CAAAGAGACC
301 AGTAAAAAAG GTATGGCAGC ATGAATAAAA AAGTTTGTAG AGTATTCATT
351 ATGCTTTCCA AATAAAAAAC TCTTTGGTTC ATAATTGTGT CATAAATTAA
401 GGACTGGGTA CACTGTACTA TTTAAAAATG TTAAGAAACA TCAATAAGTA
451 AAAATGTTAG GAAGAGATGA TAAATACGTA AGTATTATAT CTAACTAAGT
501 CTTTACTAAC TAGTCACATT ATTAACAGT GCAAGGATCA AGAAAAGTTA
551 AGCGTTGAAA AATAAATAAA TAAGTTATAA ATAAATAAAA CAGCCCAAGG
601 AAATGTTCCA GTCCCATAG GTAGACTCGG GGTCTCTTC TTTATTTAAA
651 TCTTTATTTA AATGTGGATA GCATCCCAAG AGACTTGGGT CTACACTAAG
701 AATATTCAAA TCCATGTTTC TGAAACCATC AGAGATAGAA AAAAAAAGTA
751 GCGAATATCC CTTTCAACT GGAATAAACT TGTCTTAATT CTAGAACTTT
801 TCCATACCAA TGTTCATG CTTCCTTTGT ATTTATCTTT TTAGCTCATT
851 ATCAAATTAT AGTGATTGTA AGAAGAGATC TGCTGTGAAC CTAAATGCTC
901 CTAGAACCCC AGGAAGGCAT GGATTGACAA CCACACCTCA ACAAACCTCA
951 CTCCTCCAGC ACTTGCCACA GAGGCAGGGA AATGATACAG ATAAGACTCA
1001 GGGTGACAG ACTTGTGTGG CCAACGGTGT AATGGCAGCA CAAAACCAGA
1051 TGGAATGTGA GGAGGAGAAA GCTGCCACTC TTAGCTCAGA TACTTCTATT
1101 CAAGCTTCTG AACCTTGCT TGATACGCAC ATAGTGAATG GAGAAAGAGA
1151 TGAAACTGCC ACAGCTCCTG CATCACCAC AACAGATAGC TGTGATGGAA
1201 ATGCTTCTGA CAGTAGCTAC AGGACTCCAG GCATAGGCCC AGTGCTCCCC
1251 CTAGAAGAAA GAGGGGCAGA AACAGAAACC AAGGTACAAG AGAGGGAAAA
1301 TGGGGAAAGC CCTCTGGAAC TGGAGCAGCT GGACAGCAC CATGAGATGA
1351 AGGTAGAGCA TGAGACTAGC TCATGAGCAG GGAAAACCTT GCCTATTCGA
1401 TTGTTGTCTT AAAACTCTTT ATTTATTGCA CCCCTGAAAT GTATGAATCA
1451 GATACCCAC ACTGGCAGTT AAACGATTTT CAAGCTCTGG CTGCTGATTA
1501 GCATTTCCCC TATGCTCTAA GCAGATATTT CACTTTTCTT TTTCATGTAG
1551 TTTCTGTGTA TATCTCTGTT GTAATTCAG GAGTCAGAAC AGTGTGGAAA
1601 CTTTAATATA GGAATCCAC AAATGTATTG TTTTACATA GAAAGAAAT
1651 GTTCCTTGTT GCTCTAGATG TTGGTGCTGT ATCCCTAATA CTTACGGGCC
1701 AAGCAAGAAG AAATTGTATA ATCTTTGTTG TTCAGAAGTT TCTAATAGAA
1751 TAAATAGGCC TGTAAGATGA ACTTGCCACT AGTAAATGTT ACTTTTAAGG
1801 ACATGAATAT GGAAGTATTA AATTATTCAG CAGATAAAAA AAAAA
```

BLAST Results

No BLAST result

Medline entries

98334590:
Frabin, a novel FGD1-related actin filament-binding protein capable of
changing cell shape
and activating c-Jun N-terminal kinase.

Peptide information for frame 3

ORF from 810 bp to 1373 bp; peptide length: 188
Category: similarity to known protein
Classification: Cell structure/motility

1 MFSCFLCILS FSSLSNYSYL KESAVNLNA PRTPGRHGLT TTPQKLLSQ
51 HLPQRQGNLT DKTQGAQTCV ANGVMQAQNG MECEEEKAAT LSSDTSIQAS
101 EPLLDTHIVN GERDETATAP ASPTDSCDG NASDSSYRTP GIGPVLPLEE
151 RGAETETKVQ ERENGESPLE LEQLDQHHEM KVEHETSS

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_72k15, frame 3

TREMBL:AF038388_1 product: "actin-filament binding protein Frabin";
Rattus norvegicus actin-filament binding protein Frabin mRNA, complete
cds., N = 1, Score = 428, P = 1.8e-39

>TREMBL:AF038388_1 product: "actin-filament binding protein Frabin"; Rattus
norvegicus actin-filament binding protein Frabin mRNA, complete cds.
Length = 766

HSPs:

Score = 428 (64.2 bits), Expect = 1.8e-39, P = 1.8e-39
Identities = 90/174 (51%), Positives = 115/174 (66%)

Query: 12 SLSNYSYDLKKESAVNLNAPRTTPGRHGLTTTPQKLLSQHLPQRQGNLTKTQGAQTCVA 71
S LS+Y+D++K+S +NLN P+TP +HGLT+T QKL S PQ+Q D+D+ QG C+A
Sbjct: 31 SVLSSYTDVQKDSMTNINIPQTPRQHGLTSTTPQKLP SHKSPQKQEKDSDQNQGQHGCLA 90
Query: 72 NGVMAAQNMCECEEEKAATLSSDTSIQASEPLLDTHIVNGERDETATAPASPTDSCDGN 131
NGV AAQ+QMECE EK A LS +T Q + D H++NG R+ET T AS T+S D N
Sbjct: 91 NGVAAAQSQMECEKEAALSPETDTQTAAASPDHVLNGVRNETTDSASSVTNSHDEN 150
Query: 132 ASDSSYRTPGIGPVLPLEERGAETETKVQERENGESPLELEQLDQHHEMKVEHE 185
A DSS RT G LP +E E ++QERENG S L LDQHHE+K +E
Sbjct: 151 ACDSSCRTQGTDLGLPSKEGEPVIEAELQERENGLSTEGLNPLDQHHEVKETNE 204

Pedant information for DKFZphtes3_72k15, frame 3

Report for DKFZphtes3_72k15.3

[LENGTH] 188
[MW] 20388.32
[pI] 4.62
[HOMOL] TREMBL:AF038388_1 product: "actin-filament binding protein Frabin"; Rattus
norvegicus actin-filament binding protein Frabin mRNA, complete cds. 2e-38
[KW] All Alpha
[KW] SIGNAL_PEPTIDE 16
[KW] LOW_COMPLEXITY 12.77 %

SEQ MFSCFLCILSFSSLSNYSYDLKKESAVNLNAPRTTPGRHGLTTTPQKLLSQHLPQRQGNLT
SEG .XX
PRD ccchhhhhccchhhhhhhcccccccccc
SEQ DKTQGAQTCVANGVMQAQNMCECEEEKAATLSSDTSIQASEPLLDTHIVNGERDETATAP

```
SEG .....xxxxx
PRD cccccceecchhhhhhhhhhhhhhhhhhhhhccccceecccccceeccccccccccc

SEQ ASPTTDSCDGNASDSSYRTPGIGPVLPLEERGAETETKVQERENGESPLELEQLDQHHEM
SEG xxxxx.....
PRD cccccccccccccccccccccccccccccccccchhhhhhhhhccccchhhhhhhhhhh

SEQ KVEHETSS
SEG .....
PRD hhhhcccc
```

(No Prosite data available for DKFZphtes3_72k15.3)

(No Pfam data available for DKFZphtes3_72k15.3)

DKFZphtes3_72p16

group: intracellular transport and trafficking

DKFZphtes3_72p16 encodes a novel 796 amino acid protein with very strong similarity to Mus musculus maternal-embryonic 3 (Mem3) gene.

Mem3 was isolated from a partial subtraction library of mouse unfertilized eggs and preimplantation embryos. Its transcript is abundant in the unfertilized egg and also actively transcribed from the newly formed zygotic genome. As Mem3, the novel protein is similar to yeast VPS (vacuolar protein sorting) 35. The null allele of VPS35 results in yeast in a differential defect in the sorting of vacuolar carboxypeptidase Y (CPY), proteinase A (PrA), proteinase B (PrB), and alkaline phosphatase (ALP).

The new protein can find application in modulation the sorting of proteins into different compartments.

strong similarity to mouse MEM3 and yeast VPS35

Sequenced by DKFZ

Locus: /map="16p13.3"

Insert length: 2707 bp

Poly A stretch at pos. 2697, no polyadenylation signal found

```

1 CTACGCGCGG GCGGGGTGCT GCTTGCTGCA GGCTCTGGGG AGTCGCCATG
51 CCTACAACAC AGCAGTCCCC TCAGGATGAG CAGGAAAAGC TCTTGGATGA
101 AGCCATACAG GCTGTGAAGG TCCAGTCATT CCAAATGAAG AGATGCCTGG
151 ACAAAAACAA GCTTATGGAT TCTCTAAAAC ATGCTTCTAA TATGCTTGGT
201 GAACTCCGGA CTCTATGTT ATCACCAAAG AGTTACTATG AACTTTATAT
251 GGGCATTCTT GATGAAGTGC ACTACTTGGG GGTCTACCTG ACAGATGAGT
301 TTGCTAAAGG AAGGAAAGTG GCAGATCTCT ACGAAGTGTG ACAGTATGCT
351 GGAACACATTA TCCCAAGGCT TTACCTTTTG ATCACAGTTG GAGTTGTATA
401 TGTCAAAGTCA TTTCCTCAGT CCAGGAAGGA TATTTTGAAA GATTGGTAG
451 AAATGTGCCG TGGTGTGCAA CATCCCTTGA GGGGTCTGTT TCTTCGAAAT
501 TACCTTCTTC AGTGTACCAG AAATATCTTA CCTGATGAAG GAGAGCCAAC
551 AGATGAAGAA ACAACTGGTG ACATCAGTGA TTCCATGGAT TTTGTACTGC
601 TCAACTTTGC AGAAATGAAC AAGCTCTGGG TGCGAATGCA GCATCAGGGA
651 CATAGCCGAG ATAGAGAAAA AAGAGAACGA GAAAGACAAG AACTGAGAAT
701 TTTAGTGGGA ACAAATTGGG TGCCTTCAG TCAGTTGGAA GGTGTAATG
751 TGAACAGTTA CAAACAGATT GTTTTGACTG GCATATTGGA GCAAGTTGTA
801 AACTGTAGGG ATGCTTTGGC TCAAGAAATAT CTCATGGAGT GTATTATTCA
851 GGTTCCTCCT GATGAATTC ACCTCCAGAC TTTGAATCCT TTTCTCGGG
901 CCTGTGCTGA GTTACACCAG AATGTAATG TGAAGAATC AATCATTGCT
951 TTAATTGATA GATTAGCTTT ATTGCTCAC CGTGAAGATG GACCTGGAAT
1001 CCCAGCGGAT ATTAACTTT TTGATATATT TTCACAGCAG GTGGCTACAG
1051 TGATACAGTC TAGACAAGAC ATGCCTTCAG AGGATGTTGT ATCTTTACAA
1101 GTCTCTCTGA TTAATCTTGC CATGAAATGT TACCCTGATC GTGTGGACTA
1151 TGTGTATAAA GTTCTAGAAA CAACAGTGGG GATATTCAAT AAGCTCAACC
1201 TTGAACATAT TGCTACCAGT AGTGCAGTTT CAAAGGAACT CACCAGACTT
1251 TTGAAAAATAC CAGTTGACAC TTACAACAAT ATTTTAACAG TCTTGAAATT
1301 AAAACATTTT CACCCACTCT TTGAGTACTT TGACTACGAG TCCAGAAAGA
1351 GCATGAGTTG TTATGTGCTT AGTAATGTTT TGGATTATAA CACAGAAATT
1401 GTCTCTCAAG ACCAGGTGGA TTCCATAATG AATTGGTAT CCACGTTGAT
1451 TCAAGATCAG CCAGATCAAC CTGTAGAAGA CCCTGATCCA GAAGATTTTG
1501 CTGATGAGCA GAGCCTTGTG GGCCGCTTCA TTCATCTGCT GCGCTCTGAG
1551 GACCCTGACC AGCAGTACTT GATTTTGAAC ACAGCACGAA AACATTTTGG
1601 AGCTGGTGGA AATCAGCGGA TTCGCTTCAC ACTGCCACCT TTGGTATTTG
1651 CAGCTTACCA GCTGGCTTTT CGATATAAAG AGAATTCTAA AGTGGATGAC
1701 AAATGGGAAA AGAAATGCCA GAAGATTTTT TCATTTGCCC ACCAGACTAT
1751 CAGTGCTTTG ATCAAAGCAG AGCTGGCAGA ATTGCCCTTA AGACTTTTTC
1801 TTCAAGGAGC ACTAGCTGCT GGGGAAATTG GTTTTGAAAA TCATGAGACA
1851 GTCGCATATG AATTCATGTC CCAGGCATTT TCTCTGTATG AAGATGAAAT
1901 CAGCGATTCC AAAGCACAGC TAGCTGCCAT CACCTTGATC ATTGGCACTT
1951 TTGAAGAGAT GAAGTGCTTC AGTGAAGAGA ATCATGAACC TCTGAGGACT
2001 CAGTGTGCCC TTGCTGCATC CAAACTTCTA AAGAAACCTG ATCAGGGCCG
2051 AGCTGTGAGC ACCTGTGCAC ATCTCTCTG GTCTGGCAGA AACACGGACA
2101 AAAATGGGGA GGAGCTTCAC GGAGGCAAGA GGGTAATGGA GTGCCATAAA
2151 AAAGCTCTAA AAATAGCAAA TCAGTGCATG GACCCCTCTC TACAAGTGCA
2201 GCTTTTATA GAAATTCTGA ACAGATATAT CTATTTTAT GAAAGGAAA
2251 ATGATGCGGT AACAAATCAG GTTTTAAACC AGCTTATCCA AAAGATTCCA
2301 GAAGACCTCC CGAATCTTGA ATCCAGTGAA GAAACAGAGC AGATTAACAA
2351 ACATTTTCAT AACACACTGG AGCATTTGCG CTTGCGGCGG GAATCACCAG
2401 AATCCGAGGG GCCAATTAT GAAGGTCTCA TCCTTTAAAA AGGAAATAGC
2451 TCACCATACT CTTTCCATG TACATCCAGT GAGGGTTTAA TTACGCTAGG
2501 TTTCCCTTCC ATAGATTGTG CCTTTCAGAA ATGCTGAGGT AGGTTTCCCA

```


2551 TTTCTTACCT GTGATGTGTT TTACCCAGCA CCTCCGGACA CTCACCTTCA
 2601 GGACCTTAAT AAAATTATTC ACTTGGTAAG TGTCAAGTC TTTCTGATCA
 2651 CCCCAAGTAG CATGACTGAT CTGCAATTTA AAATTCCTGT GATCTGTAAA
 2701 AAAAAA

BLAST Results

Entry AC007225 from database EMBLNEW:
 Homo sapiens chromosome 16 clone 480G7, WORKING DRAFT SEQUENCE, 38
 unordered pieces.
 Score = 1081, P = 2.8e-217, identities = 219/221
 13 exons

Entry HS015146 from database EMBL:
 human STS WI-8848.
 Score = 2033, P = 2.9e-87, identities = 425/436

Medline entries

96327632:
 Genetic mapping and embryonic expression of a novel, maternally
 transcribed gene Mem3.

97258867:
 Endosome to Golgi retrieval of the vacuolar protein sorting receptor,
 Vps10p, requires the function of the
 VPS29, VPS30, and VPS35 gene products.

92360909:
 Alternative pathways for the sorting of soluble vacuolar proteins in
 yeast: a vps35 null mutant missorts and
 secretes only a subset of vacuolar hydrolases.

10198044:
 Distinct Domains within Vps35p Mediate the Retrieval of Two Different
 Cargo Proteins from the Yeast
 Prevacuolar/Endosomal Compartment

Peptide information for frame 3

ORF from 48 bp to 2435 bp; peptide length: 796
 Category: strong similarity to known protein
 Classification: unset

1 MPTTQQSPQD EQEKLLDEAI QAVKVQSFQM KRCLDKNKLML DSLKHASNML
 51 GELRTSMLSP KSYELYMAI SDELHYLEVY LTDEFAGGRK VADLYELVQY
 101 AGNIIPRLYL LITVGVVYVK SFPQSRKDIL KDLVEMCRGV QHPLRGLFLR
 151 NYLLQCTRNI LPDEGEPTDE ETTGDISDSM DFVLLNFAEM NKLWVRMQHQ
 201 GHSRDREKRE RERQELRILV GTNLVRLSQL EGVNVERYKQ IVLTGILEQV
 251 VNCRDALAQE YLMECIIQVF PDEFHLQTLN PFLRACAEHL QNVNVKNIII
 301 ALIDRLALFA HREDGPGIPA DIKLFDFISQ QVATVIQSRQ DMPSEDVVS
 351 QVSLINLAMK CYPDRVDYVD KVLETTVEIF NKLNLHIAT SSAVSKELTR
 401 LLKIPVDYTN NILTVLKLKH FHPLFEYFDY ESRKSMSCYV LSNVLDYNTE
 451 IVSQDQVDSI MNLVSTLIQD QPDQPVDPED PEDFADEQSL VGRFIHLRLS
 501 EDPDQQYLIL NTARKHFGAG GNQRIRFTLP PLVFAAYQLA FRYKENSVD
 551 DKWEKKCQKI FSFAHQTISA LIKAEALAEPL LRLFLQGALA AGEIGFENHE
 601 TVAYEFMSQA FSLYEDEISD SKAQLAAITL IIGTFERMKC FSEENHEPLR
 651 TQCALAASKL LKKPDQGRAV STCAHLFWSG RNTDKNGEEL HGGKRVMECL
 701 KKALKIANQC MDPSLQVQLF IEILNRYIYF YEKENDAVTI QVLNQLIQKI
 751 REDLPNLESS EETEQINKHF HNTLEHLRLR RESPESEGPI YEGLIIL

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_72p16, frame 3

TREMBL:AF024504 3 gene: "A TM017A05.7"; Arabidopsis thaliana BAC
 TM017A05., N = 2, Score = 927, P = 1.9e-162

PIR:S56936 vacuolar protein-sorting protein VPS35 - yeast
(Saccharomyces cerevisiae), N = 3, Score = 826, P = 1.5e-116

TREMBL:MM47024_1 gene: "Mem3"; product: "MEM3"; Mus musculus
maternal-embryonic 3 (Mem3) mRNA, complete cds., N = 1, Score = 3376, P
= 0

TREMBL:S42186_1 gene: "VPS35"; product: "Vps35p"; VPS35=vacuolar
protein sorting [Saccharomyces cerevisiae=yeast, Genomic, 3790 nt], N =
3, Score = 813, P = 4.4e-115

>TREMBL:MM47024_1 gene: "Mem3"; product: "MEM3"; Mus musculus
maternal-embryonic 3 (Mem3) mRNA, complete cds.
Length = 754

HSPs:

Score = 3376 (506.5 bits), Expect = 0.0e+00, P = 0.0e+00
Identities = 666/721 (92%), Positives = 682/721 (94%)

```
Query:      78 EVYLTDEFAKGRKVADLYELVQYAGNIIPRLYLLITVGVVYVKSFPQSRKDILKDLVEMC 137
             +VYLTDEFAKG ++ADLYELVQY+GNIIPRLYLLITVGVVYVKSFPQSRKDILKDLVEMC
Sbjct:      34 KVIYLTDEFAKGERLADLYELVQYSGNIIPRLYLLITVGVVYVKSFPQSRKDILKDLVEMC 93

Query:     138 RGVOHPLRGLFLRNYLLQCTRNLDPDEGEPTDEETGDISDSMDVLLNFAEMNKLWVRM 197
             RGVOHPLRGLFLRNYLLQCTRNLDPDEGEPTDEETGDISDSMDVLLNFAEMNKLWVRM
Sbjct:      94 RGVOHPLRGLFLRNYLLQCTRNLDPDEGEPTDEETGDISDSMDVLLNFAEMNKLWVRM 153

Query:     198 QHQGHSRDREKRERERQELRILVGTNLVRLSOLG-VNVERYKQIVLTGILEQVVCNCRDA 256
             QHQGHSRDREKRERERQELRILVGTNLV L+ + +QIVLTGILEQVVCNCRDA
Sbjct:     154 QHQGHSRDREKRERERQELRILVGTNLVALTLVSWRCKGTLQIVLTGILEQVVCNCRDA 213

Query:     257 LAQEYLMECIIQVFPDEFHLQTLNPFRLACAEHLQNVNVKNIIALIDRLALFAHREDGP 316
             LAQE MECIIQVFPDEFHLQTLNPFRLACAEHLQNVNVKNIIALIDRLALFAHRE P
Sbjct:     214 LAQEISMECIIQVFPDEFHLQTLNPFRLACAEHLQNVNVKNIIALIDRLALFAHREMEP 273

Query:     317 GIPADIKLFDIFSQQVATVIQSRQDMPSEDVVSQVSLINLAMKCYPRVDYVDKVLETT 376
             GIPA++KLFDIFSQQVATVIQSR+DMPSEDVVSQVSLINLAMKCYPRVDYVDKVLETT
Sbjct:     274 GIPAEKLKLDIFSQQVATVIQSRDMPSEDVVSQVSLINLAMKCYPRVDYVDKVLETT 333

Query:     377 VEIFNKLNLHIATSSAVSKELTRLKIPVDTYNNILTVLKLKHFHPLFEYFDYESR--K 434
             VEIFNKLNLHIATSSAVSKELTRLKIPVDTYNNILTVLKLKHFHPLFEYFDYES K
Sbjct:     334 VEIFNKLNLHIATSSAVSKELTRLKIPVDTYNNILTVLKLKHFHPLFEYFDYESSPGK 393

Query:     435 SMSCYVLSNVLDYNTIEIVSQDQVDSIMNLVSTLIQDQDPVEDPDPEDFADEQSLVGRF 494
             SMSCYVLSNVLDYNTIEIVSQDQVDSIMNLVSTLIQDQDPVEDPDPEDFADEQSLVGRF
Sbjct:     394 SMSCYVLSNVLDYNTIEIVSQDQVDSIMNLVSTLIQDQDPVEDPDPEDFADEQSLVGRF 453

Query:     495 IHLRSDDPDQQYLIILNTARKHFGAGGNQIRFTLPPLVFAAYQLAFRYKENSKVDDKWE 554
             IHLRS+DPDQQYLIILNTARKHFGAGGNQIRFTLPPLVFAAYQLAFRYKENS K
Sbjct:     454 IHLRSDDPDQQYLIILNTARKHFGAGGNQIRFTLPPLVFAAYQLAFRYKENSQWMTSGK 513

Query:     555 KKQKIFSFHQITISALIKAEALPLRLFLQGALAAGEIGFENHETVAYEFMSQAFSLY 614
             + ++ F HQTISALIKAEALPLRLFLQGALAAGEIGFENHETVAYEFMSQAFSLY
Sbjct:     514 RNARRYFHLPHQTISALIKAEALPLRLFLQGALAAGEIGFENHETVAYEFMSQAFSLY 573

Query:     615 EDEISDSKAQLAAITLIIGTFERMKCFSEENHEPLRTQCALAASKLLKKPDQGRAVSTCA 674
             EDEISDSKAQLAAITLIIGTFERMKCFSEENHEPLRT+CALAASKLLKKPDQ C
Sbjct:     574 EDEISDSKAQLAAITLIIGTFERMKCFSEENHEPLRTECALAASKLLKKPDQAEREHMCT 633

Query:     675 HLFWSGRNTDKNGEELHGGKRVMECLKKALKIANQCMDSLQVQLFIEILNRYIYFYEKE 734
             L WSGRNTDKNGEELHGGKRVMECLKKALKIANQCMDSLQVQLFIEILNRYIYFYEKE
Sbjct:     634 SL-WSGRNTDKNGEELHGGKRVMECLKKALKIANQCMDSLQVQLFIEILNRYIYFYEKE 692

Query:     735 NDAVTIQVLNQLIQKIREDLNLESSEETEIQINKHFHNTLEHLRLRRESPESSEGPYIEGL 794
             NDAVTIQVLNQLIQKIREDLNLESSEETEIQINKHFHNTLEHLR RRESPESSEGPYIEGL
Sbjct:     693 NDAVTIQVLNQLIQKIREDLNLESSEETEIQINKHFHNTLEHLRTRRESPESSEGPYIEGL 752

Query:     795 IL 796
             IL
Sbjct:     753 IL 754
```

Pedant information for DKFZphtes3_72p16, frame 3

Report for DKFZphtes3_72p16.3

[LENGTH] 796

DKFZphtes3_7b22

group: cell structure and motility

DKFZphtes3_7b22 encodes a novel 443 amino acid protein with weak similarity to paramyosins.

The novel protein is related to paramyosin, a major structural component of thick filaments and invertebrate muscle. Paramyosins are promising antigens for immunization against several parasites, such as *Schistosoma mansoni*.

The new protein can find application in modulating cell adhesion/motility and membrane/cyto skeleton structure and dynamic.

similarity to paramyosins

complete cDNA, complete cds, few EST hits

Sequenced by BMFZ

Locus: /map="3"

Insert length: 2291 bp

Poly A stretch at pos. 2241, polyadenylation signal at pos. 2213

```
1 GGAAGAAAGG CTAGCGGGCG TTGGCCGTAT GTGGGTGTCT TGAGGCAGTT
51 TTTCACTTCT TTCATTTACC AAAGTGACAT GCACCTACTA GGTGCCAGGT
101 GTTTAGACGT ACATACAACC CTCTGCAAAA TCTTTCAGTG TAGTCCTCTG
151 TATGAAAAGT TTCCAGCCAA GAATTGCCAC TGCACCTGAG ATAAGGGGGA
201 TCCTGGCCAT TAAGGAAACC TTGCCTTCGA AACTGAGCCG TGAGGAACTA
251 TACAAAATGG GAAATTGGGA CAAATCCAG TGGCTCATGA CACTAAGAAG
301 TAAAATTACG AACTCACTGA GCTGGAAGTC ATTCAACGGG AATTGAATAG
351 GTAACCTGCAC TTTTGTGAGA TTATAAATAT ACCACGGAGG GTAACGAAGC
401 TACAGAGAA TGAAGAAGA CAGCCTGAA GACTCAAACC TTCCTCCAAA
451 AGTTTGGCAT TCTGAGATGA CCGTGTCACT GACAGGCGAA CCACCTAGTA
501 CCGTAGAAGA AGAAGGAATA CCTAAAGAAA CAGACATAGA AATCATCCCA
551 GAAATCCCGG AAACCTCTAG GCCACTGTCC CTTCCAGATG TGCTGAGGAT
601 CTGGGCAGTT CTGGAGGACA CCACAGACCA GCTCTCTATT CTGAAGTACA
651 TCATGCCCGT TCAGTACGAA GGGAGACAGA GCATCTGCGT GAAAAGCAGA
701 GAAATGAATC TAGAAGGAAC GAATCTAGAC AAACCTCCAA TGGCCTCAAC
751 AATCACAAAA ATACCCAGTC CGTTAATAAC TGAGGAAGGA CCCAACTTGC
801 CAGAAATCAG ACACAGAGGC CCGTTCGCTG TGGAGTTTAA CAAAATGCAG
851 GATCTTGTCT TCAAAAACC TACAAGGCAG ACCATCATGA CTACGGAGAC
901 ACTGAAGAAA ATTCAAGATT ATAGGCAGTT TTTCAAGCAT GTGATTGCAG
951 ATACCATTA GAGTTGCAA GATTCGGCCA CTTACAACAG TCTCCTGCAA
1001 GCTTTGAGCA AAGAGAGGGA AAACAAAATG CATTCTATG ACATCATTGC
1051 CAGGGAGGAA AAAGGAAGAA AACAGATAAT ATCACTTCAA AAACAGCTAA
1101 TTAATGTCAA AAAGGAATGG CAATTTGAAG TCCAGAGTCA GAATGAGTAT
1151 ATTGCTAACC TCAAGGACCA ACTGCAAGAG ATGAAGCAA AATCCAACCT
1201 GGAGAATCGC TACATGAAAA CCAATACCGA GCTGCAGATT GCCCAGACCC
1251 AGAAAAAGTG TAACAGAAAC GAGGAACCTT TGGTGGAGA GATTGAGAAA
1301 CTCAGGATGA AAACCGAAGA AGAGGCCCGG ACTCATACAG AGATTGAAAT
1351 GTTCCTTAGA AAGGAGCAGC AGAAACTTGA GGAGAGGCTG GAGTTCCTGA
1401 TGAGAAATTA CGATAAGGAC ACAGAAATGA AACAGAAATG ACTAAATGCT
1451 CTCAAAGCCA CAAAGGCCAG TGACTTAGCA CACCTTCAAG ACCTGGCAAA
1501 GATGATAAGA GAGTATGAAC AGGTCATCAT TGAAGATCGT ATAGAAAAGG
1551 AGAGGAGCAA GAAGAAGGTA AAACAGGATC TCTTGGAAAT AAAGAGCGTT
1601 ATAAAGCTCC AGGCCTGGTG GCGAGGCACT ATGATACGGA GAGAAATTGG
1651 TGGTTTCAAG ATGCCTAAAG ACAAAGTTGA TAGCAAGGAT TCAAAAGGCA
1701 AAGGTAAGG CAAGGATAAG AGGAGAGGCA AGAAGAAGTG ACCAAGTTCT
1751 CTTTTGTGTT TTCTGCTGGT ATTCTGGAGG TGGGAAGGAC TTGGAGAGTT
1801 AAGAAACACC TGGTACCTCA AAGATGACTC ATCTACAGGT TGTTTCTTAT
1851 TGAGACTTTC CCAGGGAAGC CTGATTTCAC TTTGCCTGTT AATTTCACCT
1901 TGCCTGTTAG GTGGGTTTTC AAACCCGTAT TTAGGATTAC ACCATTGACT
1951 TAGGGCTTCC TCATACCTTG CTGGGAAGAA GTTCTTAGTA GTCCTGTGAA
2001 GATTCACTCT TCTTGCTCTT TCTCAGCAGA ACAAAGGAGT TCACTGGCTT
2051 AGCTACAGTG ACGCATTGAA ACTTGAGTAA TTCTGTAAAT GTCAGATTTT
2101 GATTTTACCC AATTTGTCTG TAGTGAAAAA ACTCTTATGA GCAAAAGTAT
2151 TCAGTAGGAA TTACAATATG ATGTTATTAG CTGTCCAGCA TAATATATAC
2201 ACAGCAAAGT TTTAATAAAT GTTGGTTCCT GCCTGCCTTT TAAAAAATAA
2251 AAAAAAATAA AAAAAAATAA AAAAAAATAA AAAAAAATAA A
```

BLAST Results

Entry G36731 from database EMBL:
SHGC-52923 Human Homo sapiens STS cDNA.

Score = 2262, P = 1.3e-97, identities = 462/468

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 410 bp to 1738 bp; peptide length: 443
Category: similarity to known protein

```

1 MEEDSLEDSN LPPKVHSEM TVSVTGEPPS TVEEGIPKE TDIEIPEIP
51 ETLEPLSLPD VLRISAVLED TTDQLSILNY IMPVOYEGRO SICVKSREM
101 LEGTNLKLKP MASTITKIPS FLITEEGPNL PEIRHRGRFA VEFNMQDLV
151 FKKPTRQTIM TTETLKKIQI DRQFFSDVIA DTIKELQDSA TYNLLQALS
201 KERENKMHFY DIIAREEKGR KQIISLQKQL INVKKWEQFE VQSNEYIAN
251 LKQQLQEMKA KSNLENRYMK TNELQIAQT QKKCNRTTEL LVEEIEKLRM
301 KTEEEARTHT EIEMFLRKEQ QKLEERLEFW MEKYDKDTEM QNELNALK
351 TKASDLAHLQ DLAKMIREYE QVIEDRIEK ERSKKVKQD LLELKSIVKL
401 QAWWRGTMR REIGGFKMPK DKVDSKDSKG KKGKDKRRG KKK

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFzptes3_7b22, frame 2

SWISSPROT:MYSP_BRUMA PARAMYOSIN., N = 1, Score = 158, P = 5.8e-08

PIR:A44972 paramyosin - nematode (*Dirofilaria immitis*) (fragment), N = 1, Score = 157, P = 7.1e-08

SWISSPROT:MYSP_ONCVO PARAMYOSIN., N = 1, Score = 157, P = 7.4e-08

PIR:S52537 emm L 15 protein - *Streptococcus pyogenes*, N = 1, Score = 151, P = 8.6e-08

>SWISSPROT:MYSP_BRUMA PARAMYOSIN.
Length = 880

HSPs:

Score = 158 (23.7 bits), Expect = 5.8e-08, P = 5.8e-08
Identities = 66/259 (25%), Positives = 125/259 (48%)

Query: 142 EFNMQDLVFKKPTRQTIMTTETLKKIQIDRQFFSDVIADTIKELQDSATYNLLQALSK 201
+ K + L K R T E K++ + +D +A + LQ A N LL+ +
Sbjct: 169 QLKKDKHLAEKAAERFEAQTVELSNKVEDLNRHVND-LAQQRQLQ--AENNDLLKEIHD 225

Query: 202 ER---ENKMHF-YDIIAREEKGRKQIISLQKQLINVKKWEQFEVQSNEYIANLKQDLQE 257
++ +N H Y + + E+ R+++ +++ ++ + +VQ + + + D+ E
Sbjct: 226 QKVQLDNLQHVKYQLAQQLLEEARRRLEDAERERSQLQAQLH-QVQLELDSVRTALDE--E 282

Query: 258 MKAKSNLENRYMKTNELQIAQTQKKCNRTTEELLVEEIEKLRMKT-EEEARHTTEIEMFL 316
A++ E++ NTE I Q + K + L EE+E LR K +++A +IE+ L
Sbjct: 283 SAARAEAEHKLALANTE--ITQWKSFKDAEVALHHEEVEDLRKKMLQKQAEYEEQIEIML 340

Query: 317 RKEQQ--KLEERLEFWMEKYDKDTEMKQNELNALKATKASDLAHLQDLAKMIREYEQVII 374
+K Q K + RL+ +E D E QN + L+ K + L K + E + I
Sbjct: 341 QKISQLEKAKSRLQSEVEVLIVDLEKAQNTIAILERAK-----EQLEKTVNELKVIRID 393

Query: 375 EDRIEKERSKKVKQDLLELKSIVKL 400
E +E E +++ + L EL+ + L
Sbjct: 394 ELTVELEAAQREARAALAEQLKLN 419

Score = 118 (17.7 bits), Expect = 1.3e-03, P = 1.3e-03
Identities = 54/231 (23%), Positives = 108/231 (46%)

Query: 181 DTIKELQDSATYNLLQ---ALSKERENKMHFYDIIAREEKGR-KQIISLQKQLINVKK 235
D +KE+ D LQ L+++ E + RE + Q+ +Q +L +V+
Sbjct: 218 DLLKEIHDQKVQLDNLQHVKYQLAQQLLEEARRRLEDAERERSQLQAQLHQVQLELDSVRT 277

Query: 236 EWQFE--VQSQNEY-IANLKDQLQEMKAKSNLENRYMKTNTE-LQIAQTQKKCNRTTEELL 291
 E +++ E+ +A ++ + K+K + E E L+ QK+ E++
 Sbjct: 278 ALDEESAARAEAEHKLALANTEITQWKSFKDAEVALHHEEVEDLRKKMLQKQAEYEEQIE 337

Query: 292 VEEIEKLRMKTEEEARTHTEIEMF---LRKEQQKLE--ERLEFWMKEYDKDTEMKQNELN 346
 + ++K+ + ++R +E+E+ L K Q + ER + +EK + +++ +EL
 Sbjct: 338 IM-LQKISQLEKAKSRLQSEVEVLIVDLEKAQNTIAILERAKEQLEKTVNELKVRIDELE 396

Query: 347 A-LKATKASDLAHLQDLAKMIREYEQVIEDRIEKERSKKKVKQDLELKSIVI 398
 L+A + A L +L K+ YE+ + E + R KK++ DL E K +
 Sbjct: 397 VELEAAQREARAALAEQLKLNLYEKAV-EQKEALARENKKLQDDLEHAEAL 448

Score = 107 (16.1 bits), Expect = 2.1e-02, P = 2.1e-02
 Identities = 49/279 (17%), Positives = 124/279 (44%)

Query: 123 ITEEGPNLPEIRHRGRFAV-EFNKMQDLVFKKPTRQTIMTTETLKKIQIDRQFFSDVIAD 181
 I E L + R A+ E K+++L K ++ + E KK+Q D + +AD
 Sbjct: 392 IDELTVELEAAQREARAALAEQLKLNLYEKAVEQKEALAREN-KKLQDDLEHAEALAD 450

Query: 182 TIKELQDSATYNLLQALSKEKENMHFYDIIAREEKGRKQ--IISLQKQLINVKKQWQF 239
 ++L + N+ L +E + + + R+ + R Q + LQ+ I +++ Q
 Sbjct: 451 ANRKLHELDLENARLAGEIRELQTLKESEAARRDAENRAQALAEQLQRIEMERRLQE 510

Query: 240 EVQSQNEYIANLKDQLQEMKAKSNLENRYMKTNTE-LQIAQTQKKCNRT-ELLVEEIEKL 298
 + + N++ ++ + A L + + E+ + + + E E+ V+ + +
 Sbjct: 511 KEEEMEALRKNMQFEIDRLTAA--LADAEARMKAETSRLLKKYQAEIAELEMTVDNLNRA 568

Query: 299 RMKTEEEARTHTEIEMFLRKEQQKLEERLEFWMKEYDKDTEMKQNELNALKATKASDLAH 358
 ++ ++ + +E L+ + + +L+ +++Y + Q +++AL A + +
 Sbjct: 569 NIEAQTIKKQSEQLKILQASLEDTORQLQQTLDQY----ALAQKVSALSA-ELEECKV 623

Query: 359 LQDLAKMIREYEQVIEDRIEKERSKKKVKQDLELKSIVIKLQ 401
 D A R+ ++ +E+ + V +L +K+ ++ +
 Sbjct: 624 ALDNAIRARKQAEIDLEEANGRITDLVSVNNNLTAIKNKLETE 666

Pedant information for DKFZphtes3_7b22, frame 2

Report for DKFZphtes3_7b22.2

[LENGTH] 443
 [MW] 51917.95
 [PI] 6.18
 [HOMOL] PIR:S28589 trichohyalin - rabbit 2e-08
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDL058w] 7e-07
 [FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL058w] 7e-07
 [FUNCAT] 1 genome replication, transcription, recombination and repair [M. jannaschii, MJ1322] 5e-06
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YPR141c] 1e-05
 [FUNCAT] 03.13 meiosis [S. cerevisiae, YPR141c] 1e-05
 [FUNCAT] 11.01 stress response [S. cerevisiae, YPR141c] 1e-05
 [FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YPR141c] 1e-05
 [FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YPR141c] 1e-05
 [FUNCAT] 09.10 nuclear biogenesis [S. cerevisiae, YPR141c] 1e-05
 [FUNCAT] 30.05 organization of centrosome [S. cerevisiae, YPR141c] 1e-05
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YPR141c] 1e-05
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YOR216c] 3e-05
 [FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YKR095w] 6e-05
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKR095w] 6e-05
 [FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YER008c] 1e-04
 [FUNCAT] 08.16 extracellular transport [S. cerevisiae, YER008c] 1e-04
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YER008c] 1e-04
 [FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YDR356w] 2e-04
 [FUNCAT] 08.01 nuclear transport [S. cerevisiae, YDL207w] 4e-04
 [FUNCAT] 04.07 rna transport [S. cerevisiae, YDL207w] 4e-04
 [FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation, palmitoylation, farnesylation and processing) [S. cerevisiae, YKL201c] 5e-04
 [EC] 3.6.1.32 Myosin ATPase 3e-08
 [PIRKW] phosphotransferase 6e-06
 [PIRKW] citrulline 8e-06
 [PIRKW] tandem repeat 1e-07
 [PIRKW] heart 6e-06
 [PIRKW] polymorphism 4e-06
 [PIRKW] serine/threonine-specific protein kinase 6e-06
 [PIRKW] DNA binding 8e-08

[PIRKW] muscle contraction 1e-07
 [PIRKW] actin binding 3e-08
 [PIRKW] ATP 3e-08
 [PIRKW] thick filament 1e-07
 [PIRKW] phosphoprotein 3e-08
 [PIRKW] glycoprotein 4e-06
 [PIRKW] skeletal muscle 1e-07
 [PIRKW] calcium binding 8e-06
 [PIRKW] alternative splicing 3e-08
 [PIRKW] coiled coil 3e-08
 [PIRKW] P-loop 3e-08
 [PIRKW] heptad repeat 4e-06
 [PIRKW] methylated amino acid 3e-08
 [PIRKW] basement membrane 4e-06
 [PIRKW] cardiac muscle 6e-06
 [PIRKW] extracellular matrix 4e-06
 [PIRKW] hydrolase 3e-08
 [PIRKW] membrane protein 4e-06
 [PIRKW] EF hand 8e-06
 [PIRKW] cytoskeleton 8e-06
 [PIRKW] hair 8e-06
 [SUPFAM] myosin heavy chain 3e-08
 [SUPFAM] unassigned Ser/Thr or Tyr-specific protein kinases 6e-06
 [SUPFAM] calmodulin repeat homology 8e-06
 [SUPFAM] myosin motor domain homology 3e-08
 [SUPFAM] trichohyalin 8e-06
 [SUPFAM] protein kinase homology 6e-06
 [PROSITE] AMIDATION 2
 [PROSITE] CAMP_PHOSPHO_SITE 1
 [PROSITE] CK2_PHOSPHO_SITE 12
 [PROSITE] TYR_PHOSPHO_SITE 2
 [PROSITE] PKC_PHOSPHO_SITE 4
 [PROSITE] ASN_GLYCOSYLATION 1
 [KW] All_Alpha
 [KW] LOW_COMPLEXITY 10.61 %

SEQ MEEDSLEDNLPPKVVHSEMTVSVTGEPPSTVEEEGIPKETDIEIIPETLEPLSLPD
 SEGXXXXXXXXXXXXXXXXXXXXXXXXX.
 PRD cccccccccccccccccceeeccccccccceeeccccccccceeecccccccccccccc

 SEQ VLRLSAVLEDTTDLQSLILNYIMPVQYEGRQSIKVSREMNLEGTNLDKLPMASTITKIPS
 SEG
 PRD chhh

 SEQ PLITEEGPNLPEIRHRGRFAVEFNKMDLVFKKPTRQTIMTETLKKIQIDRQFFSDVIA
 SEG
 PRD hhh

 SEQ DTIKELQDSATYNSLLQALSKERENKMHFYDIIAREEKGRKQIISLQKQLINVKKEWQFE
 SEG
 PRD hhh

 SEQ VQSQNEYIANLKDQLQEMKAKSNLENRYMKTNTLQIAQTQKCKNRTEELLVEEIEKLRM
 SEG
 PRD hhh

 SEQ KTEEEARTHTTEIEMFLRKEQQKLEERLEFWMEKYDKDTEMKQNELNALKATKASDLAHLQ
 SEG
 PRD hhh

 SEQ DLAKMIREYEQVIIEDRIEKERSKKKVKQDLLELKSIVIKLQAWWRGTMIRREIGGFKMPK
 SEGx
 PRD hhhcccccc

 SEQ DKVDSKDSKGGKGGKDKRRGKKK
 SEG xxxxxxxxxxxxxxxxxxxxxxxxxxxxx
 PRD ccccccccccccccccccccccc

Prosites for DKFZphtes3_7b22.2

PS00001	285->289	ASN_GLYCOSYLATION	PDOC00001
PS00004	152->156	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	164->167	PKC_PHOSPHO_SITE	PDOC00005
PS00005	182->185	PKC_PHOSPHO_SITE	PDOC00005
PS00005	280->283	PKC_PHOSPHO_SITE	PDOC00005
PS00005	383->386	PKC_PHOSPHO_SITE	PDOC00005
PS00006	5->9	CK2_PHOSPHO_SITE	PDOC00006
PS00006	30->34	CK2_PHOSPHO_SITE	PDOC00006

PS00006	41->45	CK2_PHOSPHO_SITE	PDOC00006
PS00006	57->61	CK2_PHOSPHO_SITE	PDOC00006
PS00006	104->108	CK2_PHOSPHO_SITE	PDOC00006
PS00006	182->186	CK2_PHOSPHO_SITE	PDOC00006
PS00006	243->247	CK2_PHOSPHO_SITE	PDOC00006
PS00006	262->266	CK2_PHOSPHO_SITE	PDOC00006
PS00006	271->275	CK2_PHOSPHO_SITE	PDOC00006
PS00006	302->306	CK2_PHOSPHO_SITE	PDOC00006
PS00006	308->312	CK2_PHOSPHO_SITE	PDOC00006
PS00006	310->314	CK2_PHOSPHO_SITE	PDOC00006
PS00007	261->269	TYR_PHOSPHO_SITE	PDOC00007
PS00007	184->193	TYR_PHOSPHO_SITE	PDOC00007
PS00009	218->222	AMIDATION	PDOC00009
PS00009	439->443	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3_7b22.2)

DKFZphtes3_7d17

group: testes derived

DKFZphtes3_7d17 encodes a novel 633 amino acid protein with weak similarity to human KIAA0454.

Pfam predicts a TNFR/NGFR cysteine-rich region.

No informative BLAST results; No predictive prosite or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to KIAA0454

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 3608 bp

Poly A stretch at pos. 3587, polyadenylation signal at pos. 3570

```
1  GGGAA GTTAC  GGGAA GTTCC  ACCCAGCGTT  TCTCAGGCAA  TCTGAAGGCA
51  AATCCTGTTT  AGACCCAGGC  GAAGGTTCCT  GGTGACCCAG  GCTCTCACCA
101 GCCAATTGTC  CCTTGCCGTC  CTCCTGAGGG  TATCTGGAGC  TTCAGTGCTG
151 TGTGCTCTTG  GCCTCCACAC  TGGGGATGCC  ACTGACTCCC  ACTGTCCAGG
201 GCTTCCAGTG  GACTCTCCGA  GGCCTGATG  TAGAACTTC  CCCATTCCGT
251 GCACCAAGAG  CAGCCTCACA  TGGTGTGGGC  CGACATCAAG  AGCTGCCAGA
301 TCCAACAGTC  CCTGGCCCA  CCTCTTCTGC  CACAAACGTC  AGCATGGTGG
351 TATCTGCCGG  CCCTTGGTCC  GGTGAGAAGG  CAGAGATGAA  CATTCTAGAA
401 ATCAACAAGA  AATCGCGCCC  CCAGCTGGCA  GAGAACAAAC  AGCAGTTCAG
451 AAACCTCAAA  CAGAAATGTC  TTGTAAGTCA  AGTGGCCTAC  TTCCTGGCCA
501 ACCGGCAAAA  TAATTACGAC  TATGAAGACT  GCAAGACCT  CATAAAATCT
551 ATGCTGAGGG  ATGAGCGGCT  GCTCACAGAA  GAGAAGCTTG  CAGAGGAGCT
601 CGGGCAAGCT  GAGGAGCTCA  GGCAATATAA  AGTCTGGTT  CACTCTCAGG
651 AACGAGAGCT  GACCCAGTTA  AGGGAGAAGT  TACAGGAAGG  GAGAGATGCC
701 TCCCGCTCAT  TGAATCAGCA  TCTCCAGGCC  CTCCTCACTC  CGGATGAGCC
751 GGACAAC TCC  CAGGGACGGG  ACCTCCGAGA  ACAGCTGGCT  GAGGGATGTA
801 GGTGGGCACA  GCACCTCGTC  CAAAAGCTCA  GCCCAGAAAA  TGATGACGAT
851 GAGGATGAAG  ATGTTAAAGT  TGAGGAGGCT  GAGAAAGTAC  AGGAATTATA
901 TGCCCCCAGG  GAGGTGCAGA  AGGCTGAAGA  AAAGGAAGTC  CCTGAGGACT
951 CACTGGAGGA  GTGTGCCATC  ACTTGTTCAA  ATAGCCACCA  CCCTTGTGAG
1001 TCAACACGAG  CTTACGGGAA  CACCAGAATC  ACATTGAGG  AAGACCAAGT
1051 CGACTCAACT  CTCATTGACT  CATCCTCTCA  TGATGAATGG  TTGATGCTG
1101 TATGCATTAT  CCCAGAAAAT  GAAAGTGATC  ATGAGCAAGA  GGAAGAAAAA
1151 GGGCCAGTGT  CTCCAGGAA  TCTGCAGGAG  TCTGAAGAGG  AGGAAGCCCC
1201 CCAGGAGTCC  TGGGATGAAG  GTGATTGGAC  TCTCTCAATT  CCTCTGACA
1251 TGCTGCCTC  ATACCACTCT  GACAGGAGCA  CCTTTCACCT  AGTAGAGGAA
1301 CAGCAAGTCG  GCTTGGCTCT  TGACATAGGC  AGACATTGGT  GTGATCAAGT
1351 GAAAAAGGAG  GACCAAGAGG  CCACAAGTCC  CAGGCTCAGC  AGGGAGCTGC
1401 TGGATGAGAA  AGAGCCTGAA  GTCTTGCAGG  ACTCACTGGA  TAGATTTTAT
1451 TCAACTCCTT  TTGAGTACCT  GGAAGTGCCT  GACTTATGCC  AGCCCTACAG
1501 AAGTGACTTT  TACTCATTGC  AGGAACAACA  CCTTGGCTTG  GCTCTTGACT
1551 TGGACAGAAT  GAAAAAGGAC  CAAGAAGAGG  AAGAAGACCA  AGGCCACCCA
1601 TGCCCCAGGC  TCAGCAGAGA  GCTGCCGGAG  GTAGTAGAGC  CTGAGGACTT
1651 GCAGGACTCA  CTGGATAGAT  GGTATTCGAC  TCCTTTCAGT  TATCCAGAAC
1701 TGCTGATTC  ATGCCAGCCC  TACGGAAGTT  GCTTTTACTC  ATTGGAGGAA
1751 GAACACGTTG  GCTTTTCTCT  TGACGTGGAT  GAAATTGAAA  AGTACCAAGA
1801 AGGGGAAGAA  GATCAAAAGC  CACCATGCCC  CAGGCTCAAC  GAGGTGCTGA
1851 TGGAAAGCAG  AGAGCCTGAA  GTCTTGCAGG  ACTCACTGGA  TAGATGTTAT
1901 TCGACTACTT  CAACTTACTT  TCAACTACAT  GCCTCATTCC  AGCAGTACAG
1951 AAGTGCCTTT  TACTCATTTC  AGGAACAGGA  CGTCAGCTTG  GCCCTTGACG
2001 TGGACAATAG  GTTTTTTACT  TTGACAGTGA  TAAGGCACCA  CCTGGCCTTC
2051 CAGATGGGAG  TCATATTCCC  AACTAAGCA  GCCCTTACTA  AGCTGAGAGA
2101 TGTCATTGCT  GCAGGCAGGA  CCTATAGGCA  CATGTAGGTT  TGAATGAAAC
2151 TGTAGTTCCC  TTTGGAAGCC  CAGTCATAGG  ATGGGAAAGT  GGGCATGGCT
2201 CTATTCTTAT  TCTCAGACCA  TGCCAGTGGC  CACCTGTGCT  CAGTCTGAAG
2251 ACGTGGGACC  CAAGTTAGGT  GTGACACGTT  CACACGACTA  TGTAGCACAT
2301 GCCGGGAGTG  ATCTGCCAGA  CATTTCAATT  TGAACAGAT  ATCTCTGGGT
2351 AGCTACAAAG  TTCTCAGGG  GTTTCATTTT  GCAGGCATGT  CTCTGAGCTT
2401 CTATACCTGC  TCAAGGTCAG  TGTATCTTT  GTGTTTAGCT  CATCCAAAGG
2451 TGTTACCTTG  GTTTCATTGA  ACCTAACCCC  ATTCTTTGTA  TCTTCAGTGT
2501 TGTGTTGTTT  TAGCTGATCC  ATCTGTAACA  CAGGAGGGAT  CCTTGGCTGA
2551 GGATTGTATT  TCAGAACCAC  TGACTGTCTT  TGACAGTTGT  TAACCCACTA
2601 GGCTCCTTTG  AGTAGAGAAG  CCATAGTCCT  TCAGCCTCCA  ATTGATATCA
2651 ATACTTAGGA  AGACCACAGC  TAGACGGACA  AACAGCATTG  GGAGGCCTTA
```

```

2701 GTCCTGCTCC TTTCAATTCC ATCCTGTAAA GAACAGGAGT CAGGAGCCGC
2751 TGGCAAGAGA CAGCATGTCA CCTGGGACTC TGCCAGTGCA GAATATGAAC
2801 AATGCCATGT TCTTGACAGAA AATGCTTAGC CTGAGTTTCA TAGGAGGTAA
2851 TCACCAGACA ACTGCAGAA GTAGAACACT GAGCAGGACA ACTGACCTGT
2901 CTCCTTCACA CAGTCCACGT CACCACGAAT CACACAACAA AAAGGAGGAG
2951 AGATATTTTG GGTTCAGAA AAGTAAATGA TAATGTAGCT ACATTCTTT
3001 AGTTATTTTG AACCCAAAT ATTCCTCAT CTTTTGTG TGTCATTGA
3051 TTTTGGTGAC ATGGACTTGT TTGTAGAGGA CAGGTCAGCT GTCTGGCTCA
3101 ATGGTCTACA TTCTGAAGTT GTCTGAAAT GTCTTCATGA TTAAATTCAG
3151 CCTAACGTT TCATCAAGAA CACTACAGAG TCGATACTGT GAGTTTCAA
3201 CCTCAGCCCA TCTGTGGGCA GAGAAGGTCT AGTTTGTCCA TCAGCATTAT
3251 CATGATATCA GGACTGGTTA CTTGGTTAAG GAGGGGTCTA GGAGATCTGT
3301 CCCTTT TAGA GACACCTTAC TTATGATGAA GTATTGGGA GAGTGGTTT
3351 TCAAAGTAGA AATGTCCTGT ATTCCAGTGA TCATCCTCTA AACGTTTTAT
3401 CATTATATTA TCATCCCTGC CTGTGTCTAT TATTATATTC ATATCTCTAC
3451 GCTGGAATTT TGCTGCCTCA ATGTTTACTG TGCCTTTGT TTTGCTAGTG
3501 TGTGTTGTTG AAAAAAAAC ATTCTCTGCC TGAGTTTAA TTTTGTCCA
3551 AAGTTATTTT AATCTATACA ATTAATAACT TTTGCCTATC AAAAAAAAC
3601 AAAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 176 bp to 2074 bp; peptide length: 633
 Category: similarity to known protein

```

1  MPLTPTVQGF QWTLRGPDVE TSPFGAPRAA SHGVGRHQEL RDPTVPGPPTS
51  SATNVSMVVS AGPWSGEKAE MNILEINKKS RPQLAENKQQ FRNLKQKCLV
101  TQVAYFLANR QNNYDYEDCK DLIKSMRLDE RLLTEEKLA E LGQAEELRQ
151  YKVLVHSQER ELTQLREKLQ EGRDASRLN QHLQALLTPD EPDNSQGRDL
201  REQLAEGCRL AQHLVQKLSP ENDDDEDEDV KVEEAQKQVE LYAPREVQKA
251  EEKEVPEDSL EECATCSNS HHPCESNQPY GNTRITFEED QVDSTLIDSS
301  SHDEWLDAVC IIPENESDHE QEEKGPVSP RNLQSEEEEE APQESWDEGD
351  WTLISIPPM SASYQSDRSTF HSVEEQVGL ALDIGRHWCD QVKEDQEAT
401  SPRLSRELLD EKEPEVLQDS LDRFYSTPFE YLELPDLQCP YRSDFYSLQE
451  QHLGLALDLD RMKKDQEEEE DQGPPCPRLS RELPEVVEPE DLQDSLDRWY
501  STPFSTPELP DSCQPYGSCF YSLEEEHVGF SLDVDEIEKY QEGEEDQKPP
551  CPRLNEVLME AEEPEVLQDS LDRCYSTTST YFQLHASFQQ YRSFYSFEE
601  QDVSLLALDVD NRFFTLTVIR HHLAFQMGVI FPH

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_7d17, frame 2

PIR:T00069 hypothetical protein KIAA0454 - human (fragment), N = 1,
 Score = 199, P = 1e-11

PIR:A45592 liver stage antigen LSA-1 - Plasmodium falciparum, N = 1,
 Score = 158, P = 2.7e-07

>PIR:T00069 hypothetical protein KIAA0454 - human (fragment)
 Length = 1,882

HSPs:

Score = 199 (29.9 bits), Expect = 1.0e-11, P = 1.0e-11
 Identities = 74/261 (28%), Positives = 122/261 (46%)

Query: 117 EDCKDLIKSMRLRDERLLT----EEKLAEELGQAEELRQYKVLVHSQERELTQLREKLQEG 172
 +D + LI+ + + E L EEKLAEE L A +Y L+ Q REL+ LR+K++EG
 Sbjct: 964 KDLESLIQRVSQLEAQLPKNGLEEKLAEEELRSASWPQKYSLSIQDQARELSYLRQKIREG 1023

Query: 173 RDASRSLNQH-----LQALLTPDEPDNSQGRDLREQLAEGCRLAQHLVQKLSPENDDD 225
 R + +H + LL ++ D G+ REQLA+G +L + L KLS ++
 Sbjct: 1024 RGICYLITRHAKOTVKSFE DLLRSNDIDYLLGQSFREQLAQGSQLTERLTSKLSTKDHKS 1083

Query: 226 EDEDVKVEEAQVQELYAPREVQKAEK-EVPEDSLEECATCSNSHHPCESNQPYGNTR 284
 E + +E L RE+Q+ E+ EV + L+ ++T S+SH +S++ +T
 Sbjct: 1084 EKDQAGLEPLA----LRLSRELQEKEKVIEWLQAKLDARSLTPSSSHALSDSHRSPSSTS 1139

Query: 285 ITFEEDQV--DSTLIDSSSHDEWLDAVCIIPENESDHEQEEKGVPSPRNQSEEEEEAP 342
 +E + D ++ +H E A P + +S + S + A
 Sbjct: 1140 FLSDELEACSDMDIVSEYTHYEKKAS---PSHSDSIHSSSHSAVLSSKPSSTSASQGA 1196

Query: 343 QESWDEGDWTLSPDMSASYQSDRSTFH 371
 ES + +L P + S FH
 Sbjct: 1197 AES-NSNPISLPTPQNTPKANQAHS GFH 1224

Score = 89 (13.4 bits), Expect = 1.1e-01, P = 1.0e-01
 Identities = 35/89 (39%), Positives = 44/89 (49%)

Query: 464 KDQEEEDQG---PPCFRLSRELPEVVEP-EDLQDSLDRWYSTPFSYPELPDSCQ-PYGS 518
 KD + E+DQ P RLSREL E + E LQ LD TP S L DS + P +
 Sbjct: 1079 KDHKSEKQAGLEPLALRLSRELQEKEKVIEWLQAKLDARSLTPSSSHALSDSHRSPSST 1138

Query: 519 CFYSLEEEHVGFSLDVDEIEKYQEGEEDQKPP 550
 F S E E D+D + +Y EE + P
 Sbjct: 1139 SFLSDELEACS---DMDIVSEYTHYEKKASP 1167

Score = 73 (11.0 bits), Expect = 4.8e+00, P = 9.9e-01
 Identities = 31/88 (35%), Positives = 40/88 (45%)

Query: 390 DQVKKEDQEATSP---RLSRELLD-EKEPEVLQDSLDRFYSTPFEYLELPDLCQ-PYRSD 444
 D ++DQ P RLSREL + EK EVLQ LD TP L D + P +
 Sbjct: 1080 DHKSEKQAGLEPLALRLSRELQEKEKVIEWLQAKLDARSLTPSSSHALSDSHRSPSSTS 1139

Query: 445 FYSLQEQHLGLALDLDRMKKDQEEEDQGGP 475
 F S L D+D + + EE + P
 Sbjct: 1140 FLS---DELEACSDMDIVSEYTHYEKKASP 1167

Score = 68 (10.2 bits), Expect = 1.1e-01, P = 1.0e-01
 Identities = 36/156 (23%), Positives = 68/156 (43%)

Query: 31 SHGVGRHQELRDPTV---PGPTSSATNVSMVVSAGPWS-----GEKAEMNILEINKK 79
 S G +HQE + TV P P S + V A G ++ ++ +
 Sbjct: 684 SPGKHQHQEGNVTVRPFPRPQSLDLGATFTVDAHQLDNQSQPRDPGPQSAFSLPGSTQH 743

Query: 80 SRPQLAENKQQRNLKQKCLVTQVAYFL-ANRQNNYDYE-DCKDLIKSMLRDERLLTEEK 137
 R QL++ KQ++L++K L+++ F AN Y + L+K + ++ ++
 Sbjct: 744 LRSQLSQCKQRYQDLQEKLLSEATVFAQANELEKYRVMLTGESLVKQDSKQIQVDLQDL 803

Query: 138 LAEELGQAEELRQYKVLVHSQERELTQREK-LQEG 172
 E G++E + + + E L+E L EG
 Sbjct: 804 GYETCGRSENEAEREETTSPECEEHNSLKEMVLMEG 839

Score = 65 (9.8 bits), Expect = 2.2e-01, P = 2.0e-01
 Identities = 23/96 (23%), Positives = 52/96 (54%)

Query: 123 IKSMRLRDERLLTEEKLAELGQAE-----LRQYKVLVHSQERELTQREKLQEGRDASRS 178
 ++ + D+ + E + E+ EE LRQ ++ V ++ +L +LR+ L ++ +
 Sbjct: 5 LRQRIHDKAVALERAIDEKFSALEEKEKELRQLRLAVRERDHDLERLRDVL-----SNEA 60

Query: 179 LNQHLOALLTPDEPDNSQGRDLREQLAEGCRLAQHLVQKL 218
 Q +++LL ++G ++ EQL+ C+ Q L +++
 Sbjct: 61 TMQSMESLL-----RAKGLEV-EQLSTTCQNLQWLKEEM 93

Score = 61 (9.2 bits), Expect = 5.5e-01, P = 4.2e-01
 Identities = 27/95 (28%), Positives = 47/95 (49%)

Query: 134 TEEK-LAEELGQAEELRQY---KVLVHSQERELTQREKLQEGRDASRS LNQHLOALLT 188
 +E K L +LG+ EE R Y +LV +++ L+ +LQ ++L +++L
 Sbjct: 855 SERKPLENQLGKQEEFRVYKSENILV--LRKDIKDLKAQLQANKVIQNLKSRVRSLSV 912

Query: 189 PDEPDNSQGRDLREQLAEGCRLAQHLVQKLSPENDDDDEDE 228
 + +S R R+ A G ++ SP + DEDE
 Sbjct: 913 TSDYSSSLERP-RKLRAVGT-----LEGSSPHSVPEDE 945

Score = 57 (8.6 bits), Expect = 1.4e+00, P = 7.5e-01
 Identities = 26/92 (28%), Positives = 47/92 (51%)

Query: 127 LRDERLLTEEKLAELGQAEEL---RQYKVLVHSQERELTQREKLQEGRDASRS LNQH 183
 L E LL EK+A Q+E+ R+ ++L+ + L R +L E A R L L
 Sbjct: 358 LTQEVLLLREKVASVESQGEISGNRRQQLLLLEG--LVDESRSLNEALQAERQLYSSL 415

Query: 184 QALLTPDEPDNSQ-GRDLREQLAEGCRLAQHLVQKL 218
 P++S+ R L+ +L EG ++ + +++
 Sbjct: 416 VKFHA--HPESSEDRDLQVEL-EGAQVLRSLREEV 448

Score = 54 (8.1 bits), Expect = 2.7e+00, P = 9.3e-01
 Identities = 61/264 (23%), Positives = 121/264 (45%)

Query: 3 LTPTVQGFQWTLRGPDVETSPFGAPRAASHGVGRHQE--LRDPTVPGPTSSATNVSMVVS 60
 L+ T Q QW L+ ++ET F + + + L D SAT ++
 Sbjct: 79 LSTTCQLQW-LK-EEMETK-FSRWQKEQESIIQQLQTSLHNRNKEVEDLSAT---LLCK 132

Query: 61 AGPWSGEKAEMNILEINKKSR--PQLAENKQFRNLKQKCLVTQVAYFLANRQNNYDYE 117
 GP E AE + +K R L++ +Q L+ + + + ++ R+
 Sbjct: 133 LGPGQSEIAEELCQRLQKRMQLDQLSDRNKQV--LEHEMEIQGLLQSVSTREQE-SQA 189

Query: 118 DCKDLIKSMLRDERLLTEEKLAEEELQAEELRQYKVLVHSQERELT---QLREKLQEG-- 172
 + L++++ ER + L + LG + L + + +Q+ E+T +L ++ +G
 Sbjct: 190 AAEKLVQALM--ERNSELQALRQYLGGDRSLMS-QAPISNQAEVTPTRGLGKQTDQGS 246

Query: 173 RDASRSLNQHLQALLTPDEPDNSQGRDLREQLAEGCRLAQHLVQKLSPENDDDDEDEDVKV 232
 + SR + L A P ++ G DL + +A G L ++LS N +E E +
 Sbjct: 247 QIFSRDSTSLTAKEDVSIPTSLG-DL-DTVA-G-----LEKELS--NAKEELELM 295

Query: 233 EEAQVQELYAPREVQKAEKEVPEDSLEECAIT 266
 +E E EL A + + +E+ + + +T
 Sbjct: 296 KERESQMEALSALQSMMAVQEELQVQAADMESLT 329

Score = 49 (7.4 bits), Expect = 6.3e+00, P = 1.0e+00
 Identities = 21/87 (24%), Positives = 39/87 (44%)

Query: 192 PDNSQGRDLREQLAEGCRLAQHLVQKLSPENDDDDEDEDVKVEEAQVQELYAPREVQKAE 251
 P ++Q LR QL++ + Q L +KL + + E EK + + + K +
 Sbjct: 738 PGSTQ--HLRSQSQCKQRYQDLQEKLLS---EATVFAQANELEKYRVMGTGESLVKQD 792

Query: 252 EKEVPEDSLEECAI-TCSNSHPCESNQ 278
 K++ D L++ TC S + E +
 Sbjct: 793 SKQIQVD-LQDLGYETCGRSENEAEREE 819

Score = 46 (6.9 bits), Expect = 6.3e+00, P = 1.0e+00
 Identities = 19/77 (24%), Positives = 39/77 (50%)

Query: 112 NNYDYEDCKDLIKSMLRDERLLTEEKLAEEELQAEELRQYKVLVHSQERELTQLREKLQ- 170
 + ++ E+ K + E ++T+E L+E QAE R+ + + + L+E+L
 Sbjct: 597 DGWEIEEDKE--KGEVMVETVVTKEGLSESLQAE-FRKLQGKLKNAHNIINLLKEQLVL 653

Query: 171 EGRDASRSLNQHLQALLT 188
 ++ + L L LT
 Sbjct: 654 SSKEGNSKLTPELLVHLT 671

Pedant information for DKFZphtes3_7d17, frame 2

Report for DKFZphtes3_7d17.2

[LENGTH] 633
 [MW] 72951.15
 [pI] 4.40
 [HOMOL] PIR:T00069 hypothetical protein KIAA0454 - human (fragment) 2e-11
 [BLOCKS] BL00201E
 [PROSITE] MYRISTYL 2
 [PROSITE] CK2_PHOSPHO_SITE 14
 [PROSITE] PKC_PHOSPHO_SITE 4
 [PROSITE] ASN_GLYCOSYLATION 2
 [PFAM] TNFR/NGFR cysteine-rich region
 [KW] All Alpha
 [KW] LOW_COMPLEXITY 4.90 %
 [KW] COILED_COIL 6.95 %

SEQ MPLTPTVQGFQWTLRGPDVETSPFGAPRAASHGVGRHQELRDPTVPGPTSSATNVSMVVS
 SEG
 PRD ccc
 COILS

SEQ AGPWSGEKAEMNILEINKKSRPQLAENKQFRNLKQKCLVTQVAYFLANRQNNYDIEDCK
 SEG
 PRD cccccchhhhhhhheccccchhhhhhhhhccccchhhhhhhhhccccccccch
 COILS

Prosite for DKFZphtes3 7d17.2

Pfam for DKF2phtes3 7d17.2

941

DKFZphtes3_7j3

group: cell cycle

DKFZphtes3_7j3.2 encodes a novel 628 amino acid putative protein kinase, which is related to the C-TAK1 Cdc25C associated protein kinase.

Cdc25C is a protein kinase that controls entry into mitosis by dephosphorylation of Cdc2. Cdc25C function is regulated by phosphorylation, too. Serine 216 phosphorylation of Cdc25C mediates the binding of 14-3-3 protein to Cdc25C. C-TAK1 (Cdc twenty-five C associated protein kinase) phosphorylates Cdc25C on serine 216 in vitro. The new protein is closely related to C-Tak1 and therefore should be involved in cell-cycle regulation, too.

The new protein can find application in modulating/blocking the cell cycle.

strong similarity to serine/threonine-specific protein kinases

complete cDNA, complete cds, potential start at Bp 128, few EST hits

Sequenced by BMF2

Locus: unknown

Insert length: 3443 bp

Poly A stretch at pos. 3399, polyadenylation signal at pos. 3376

```

1 GTGCTTTACT GCGCGCTCTG GTACTGCTGT GGCTCCCCGT CCTGGTGCGG
51 GACCTGTGCC CCGCGCTTCA GCCCTCCCCG CACAGCCTAC TGATTCCCCT
101 GCCGCCCTTG CTCACCTCCT GCTCGCCATG GAGTCGCTGG TTTTCGCGCG
151 GCGCTCCGGC CCCACTCCCT CGGCCGCGAG GCTAGCCCGG CCGCTGGCGG
201 AAGGGCTGAT CAAGTCGCCC AAGCCCTTAA TGAAGAAGCA GGCGGTGAAG
251 CGGCACCAAC ACAAGCACAA CCTGCGGCAC CGCTACGAGT TCCTGGAGAC
301 CCTGGGCAAA GGCACCTACG GGAAGGTGAA GAAGGCGCGG GAGAGCTCGG
351 GGCGCCTGGT GGCCATCAAG TCAATCCGGA AGGACAAAAT CAAAGATGAG
401 CAAGATCTGA TGCACATACG GAGGGAGATT GAGATCATGT CATCACTCAA
451 CCACCCTCAC ATCATTGCCA TCCATGAAGT GTTTGAGAAC AGCAGCAAGA
501 TCGTGATCGT CATGGAGTAT GCCAGCCGGG GCGACCTTTA TGAATACATC
551 AGCGAGCGGC AGCAGCTCAG TGAGCGCGAA GCTAGGCATT TCTTCCGGCA
601 GATCGTCTCT GCCGTGCACT ATTGCCATCA GAACAGAGTT GTCCACCGAG
651 ATCTCAAGCT GGAGAACATC CTCTTGGATG CCAATGGGAA TATCAAGATT
701 GCTGACTTCG GCCTCTCCAA CCTCTACCAT CAAGGCAAGT TCCTCGAGAC
751 ATTCTGTGGG AGCCCCCTCT ATGCTCGGCC AGAGATTGTC AATGGGAAGC
801 CCTACACAGG CCCAGAGGTG GACAGCTGGT CCCTGGGTGT TCTCTCTTAC
851 ATCTGGTGGG ATGGCAACAT GCCCTTTGAT GGGCATGACC ATAAGATCCT
901 AGTGAACACG ATCAGCAACG GGGCTTACCG GGAGCCACCT AAACCCCTCT
951 ATGCTGTGGG CCTGATCCGG TGGCTGTTGA TGGTGAACCC CACCCGCGCG
1001 GCCACCCCTG AGGATGTGGC CAGTCACTGG TGGGTCAACT GGGGCTACGC
1051 CACCCGAGTG GGAGAGCAGG AGGCTCCGCA TGAGGGTGGG CACCCCTGGC
1101 GTGACTCTGC CCGCGCCTCC ATGGCTGACT GGCTCCGGCG TTCCTCCCGC
1151 CCCCTCCTGG AGAATGGGGC CAAGGTGTGC AGCTTCTTCA AGCAGCATGC
1201 ACCTGGTGGG GGAAGCACCA CCCCTGGCCT GGAGCGCCAG CATTGCTCTA
1251 AGAAGTCCCG CAAGGAGAAT GACATGGCCC AGTCTCTCCA CAGTGACACG
1301 GCTGATGACA CTGCCATCG CCCTGGCAAG AGCAACCTCA AGCTGCCAAA
1351 GGGCATTCTC AAGAAGAAGG TGTCAAGCTC TGCAGAAGGG GTACAGGAGG
1401 ACCCTCCGGA GCTCAGCCCA ATCCCTGCGA GCCCAGGGCA GGCTGCCCGG
1451 CTGCTCCCCA AGAAGGGCAT TCTCAAGAAG CCCCGACAGC GCGAGTCTGG
1501 CTACTACTCC TCTCCGAGC CCAGTGAATC TGGGGAGCTC TTGGACGCAG
1551 GCGACGTGTT TGTGAGTGGG GATCCCAAGG AGCAGAAGCC TCCGCAAGCT
1601 TCAGGGCTGC TCCTCCATCG CAAAGGCATC CTCAACTCA ATGGCAAGTT
1651 CTCCCAGACA GCCTTGGAGC TCGCGGCCCC CACCACCTTC GGCTCCCTGG
1701 ATGAACCTCG CCCACCTCGC CCCCTGGCCC GGGCCAGCCG ACCCTCAGGG
1751 CTGTGAGCG AGGACAGCAT CCTGTCTCTT GAGTCTTTTG ACCAGCTGGA
1801 CTTGCCTGAA CGGCTCCAG AGCCCCACT GCGGGGCTGT GTGTCTGTGG
1851 ACAACCTCAC GGGGCTTGAG GAGCCCCCTC CAGAGGGCCC TGGAAGCTGC
1901 CTGAGGCGCT GCGGCGAGGA TCCTTTGGGG GACAGCTGCT TTTCCCTGAC
1951 AGACTGCCAG GAGGTGACAG CGACCTACCG ACAGGCACTG AGGGTCTGCT
2001 CAAAGCTCAC CTGAGTGGAG TAGGCATTGC CCCAGCCCGG TCAGGCTCTC
2051 AGATGACAGT GGTTCACACC CGAGGGGAGA TGCCTTCTCC CCCACCTCCC
2101 AGGACCTGCA TCCCAGCTCA GAAGGCTGAG AGGGTTTGCA GTGGAGCCCT
2151 GAGCAGGGCT GGAATATGGA AGTAGGCAAA TGAATGCGC CAAGGGTTCA
2201 GTGTCTGTCT TCAGCCCTGC TGAACGAAGA GGATACTAAA GAGAGGGGAA
2251 CGGGAATGCC CGCGACAGAG TCCACATTGC CTGTTTCTTG TGTACATGGG
2301 GGGGCCACAG AGACCTGGAA AGAGAATCTT CCCAGGGCCC ATCTCCTGCA
2351 TCCCATGAAT ACTCTGTACA CATGGTGCCT TCTAAGGACA GCTCCTTCCC
2401 TACTCATTCC CTGCCCAAGT GGGGCCAGAC CTCTTTACAC ACACATTCCC
2451 GTTCTCTACA ACCACCAGAA CTGGATGGTG GCACCCCTAA TGTGCATGAG
2501 GCATCCTGGG AATGGTCTGG AGTAACGCTT CGTTATTTT ATTCTTATT

```

```

2551 TTATTTATTT ATTTATTTTT TTGAGACGGA GTTTCGCTCT TGGTGCCAG
2601 GCTAGAGTGC AATGGCGCGA TCTCAGCTCA CCTCAACCTC CGCCTCCCGG
2651 GTTCAAGCGA TTCTCCTGCC TCAGCCTCCC TAGTAGCTGG GATTACAGGC
2701 GCCCCGCCACC ATGCCCGGCT AATTTTGTAT TTTTAGTAGA GACAGGGTTT
2751 CTCCATGTTG GTCAGGCTGG TCTCAAACCT CCGACCTCAG GTGATCCACC
2801 CACCTCGGCC TCCCAAAGTG CTGGGATTAC AGGCGTGAGC CACCGCGCCC
2851 CACCTAACCC TTCCTTATTT AGCCTAGGAG TAAGAGAACA CAATCTCTGT
2901 TTCTTCAATG GTTCTCTTCC CTTTCCATC CTCCAAACCT GGCCTGAGCC
2951 TCCTGAAAGT GCTGCTGTGA ATCTGAAAGA CTGAAAAGC CTCGCGCTGC
3001 TGTGTGGACT TCATCTCAAG GGGCCAGCC TCCTCTGGAC TCCACCTGG
3051 ACCTCAGTGA CTCAGAACTT CTGCCTCTAA GCTGCTCTAA AGTCCAGACT
3101 ATGGATGTGT TCTCTAGGCC TTCAGGACTC TAGAATGTCC ATATTTATTT
3151 TTATGTCTT GGCTTTGTGT TTTAGGAAA GTGAATCTTG CTGTTTCAA
3201 TAATGTGAAT GCTATGTTCT GGGAAATCC ACTATGACAT CTAAGTTTGT
3251 TGTACAGAGA GATATTTTGT CAACTATTTC CACCTCTCC CACAACCCCC
3301 CACACTCCAC TCCACACTCT TGAGTCTCTT TACCTAATGG TCTCTACCTA
3351 ATGGACCTCC GTGGCCAAAA AGTACCATTA AAACCAGAAA GGTGATTGGA
3401 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAA

```

BLAST Results

No BLAST result

Medline entries

98202387:
C-TAK1 protein kinase phosphorylates human Cdc25C on serine 216 and promotes 14-3-3 protein binding.

Peptide information for frame 2

ORF from 128 bp to 2011 bp; peptide length: 628
Category: strong similarity to known protein

```

1 MESLVFARRS GPTPSAAELA RPLAELIKS PKPLMKQAV KRHHHKHNL
51 HRYEFLETLG KGTYGKVKKA RESSGRLVAI KSIRKDKIKD EQDLMHIRRE
101 IEIMSSLNHP HIIAIEHVFE NSSKIVIVME YASRGDLYDY ISERQQLSER
151 EARHFFRQIV SAVHYCHQNR VVHRDLKLEN ILLDANGNIK IADFGLSNLY
201 HQCKFLQTFC GSPLYASPEI VNGKPYTGPE VDSWSLGLVLL YILVHGTMPF
251 DGHDKILVK QISNGAYREP PKPSDACGLI RWLLMVNPTN RATLEDVASH
301 WVVNWGYATR VGEQEAPHEG GHPGSDSARA SMADWLRRSS RPLENGAKV
351 CSFFKQHAPG GGSTTPGLER QHSLKKSKE NDMAQSLHSD TADDTAHRPG
401 KSNLKLPGKI LKKKVSASAE GVQEDPEELS PIPASPGQAA PLLPKKGILK
451 KPRQRESGY SSPEPSESGE LLDAGDVFSV GDPKEQKPPQ ASGLLLHRKG
501 ILKLNKFSQ TALELAAPT FGSDELAPP RPLARASRPS GAVSEDSILS
551 SESFDQLDLP ERLPEPLRG CVSVDNLTLG EEPPEGPGS CLRRWRQDPL
601 GDSCFSLTDC QEVATYRQA LRVCCKLT

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_7j3, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphtes3_7j3, frame 2

Report for DKFZphtes3_7j3.2

```

[LENGTH]      628
[MW]           69612.39
[pI]           9.01
[HOMOL]        TREMBL:AB011109_1 gene: "KIAA0537"; product: "KIAA0537 protein"; Homo sapiens
mRNA for KIAA0537 protein, complete cds. 1e-152
[FUNCAT]       01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YDR477w]
5e-66
[FUNCAT]       11.01 stress response [S. cerevisiae, YDR477w] 5e-66

```


[FUNCAT]	30.03 organization of cytoplasm	[S. cerevisiae, YDR477w]	5e-66
[FUNCAT]	98 classification not yet clear-cut	[S. cerevisiae, YLR096w]	6e-54
[FUNCAT]	30.02 organization of plasma membrane	[S. cerevisiae, YLR096w]	6e-54
[FUNCAT]	03.04 budding, cell polarity and filament formation	[S. cerevisiae, YDR507c]	8e-52
[FUNCAT]	03.25 cytokinesis	[S. cerevisiae, YDR507c]	8e-52
[FUNCAT]	03.22 cell cycle control and mitosis	[S. cerevisiae, YKL101w]	9e-51
[FUNCAT]	30.10 nuclear organization	[S. cerevisiae, YKL101w]	9e-51
[FUNCAT]	99 unclassified proteins	[S. cerevisiae, YPL141c]	1e-45
[FUNCAT]	10.99 other signal-transduction activities	[S. cerevisiae, YPL153c]	6e-44
[FUNCAT]	03.22.01 cell cycle check point proteins	[S. cerevisiae, YPL153c]	6e-44
[FUNCAT]	11.04 dna repair (direct repair, base excision repair and nucleotide excision repair)	[S. cerevisiae, YPL153c]	6e-44
[FUNCAT]	03.19 recombination and dna repair	[S. cerevisiae, YPL153c]	6e-44
[FUNCAT]	03.16 dna synthesis and replication	[S. cerevisiae, YMR001c]	2e-42
[FUNCAT]	10.02.11 key kinases	[S. cerevisiae, YBL105c]	3e-34
[FUNCAT]	04.05.01.04 transcriptional control	[S. cerevisiae, YKL139w CTK1 - carboxy-terminal domain]	2e-28
[FUNCAT]	03.01 cell growth	[S. cerevisiae, YFR014c]	4e-28
[FUNCAT]	03.10 sporulation and germination	[S. cerevisiae, YGL180w]	2e-26
[FUNCAT]	06.13.04 lysosomal and vacuolar degradation	[S. cerevisiae, YGL180w]	2e-26
[FUNCAT]	08.13 vacuolar transport	[S. cerevisiae, YGL180w]	2e-26
[FUNCAT]	04.99 other transcription activities	[S. cerevisiae, YER129w]	4e-26
[FUNCAT]	02.19 metabolism of energy reserves (glycogen, trehalose)	[S. cerevisiae, YPL031c]	5e-24
[FUNCAT]	01.04.04 regulation of phosphate utilization	[S. cerevisiae, YPL031c]	5e-24
[FUNCAT]	03.07 pheromone response, mating-type determination, sex-specific proteins	[S. cerevisiae, YHL007c]	6e-24
[FUNCAT]	10.05.11 key kinases	[S. cerevisiae, YHL007c]	6e-24
[FUNCAT]	09.01 biogenesis of cell wall	[S. cerevisiae, YNR031c]	1e-22
[FUNCAT]	10.03.11 key kinases	[S. cerevisiae, YNR031c]	1e-22
[FUNCAT]	03.13 meiosis	[S. cerevisiae, YDR523c]	8e-22
[FUNCAT]	04.05.01.01 general transcription activities	[S. cerevisiae, YDL108w]	6e-21
[FUNCAT]	06.07 protein modification (glycosylation, acylation, myristylation, palmitoylation, farnesylation and processing)	[S. cerevisiae, YFL033c]	6e-21
[FUNCAT]	10.05.09 regulation of g-protein activity	[S. cerevisiae, YBL016w]	7e-19
[FUNCAT]	10.04.11 key kinases	[S. cerevisiae, YDL159w]	3e-18
[FUNCAT]	01.02.04 regulation of nitrogen and sulphur utilization	[S. cerevisiae, YNL183c]	1e-17
[FUNCAT]	08.99 other intracellular-transport activities	[S. cerevisiae, YNL183c]	1e-17
[FUNCAT]	05.07 translational control	[S. cerevisiae, YDR283c]	2e-17
[FUNCAT]	09.04 biogenesis of cytoskeleton	[S. cerevisiae, YNL020c]	4e-16
[FUNCAT]	04.03.99 other trna-transcription activities	[S. cerevisiae, YOR061w]	1e-15
[FUNCAT]	10.04.99 other nutritional-response activities	[S. cerevisiae, YJR059w]	5e-15
[FUNCAT]	c energy conversion	[M. genitalium, MG109]	3e-12
[FUNCAT]	30.09 organization of intracellular transport vesicles	[S. cerevisiae, YBR097w]	2e-08
[FUNCAT]	08.07 vesicular transport (golgi network, etc.)	[S. cerevisiae, YBR097w]	2e-08
[FUNCAT]	06.04 protein targeting, sorting and translocation	[S. cerevisiae, YBR097w]	2e-08
[FUNCAT]	30.08 organization of golgi	[S. cerevisiae, YBR097w]	2e-08
[FUNCAT]	30.07 organization of endoplasmatic reticulum	[S. cerevisiae, YHR079c]	8e-05
[FUNCAT]	01.06.10 regulation of lipid, fatty-acid and sterol biosynthesis	[S. cerevisiae, YHR079c]	8e-05
[BLOCKS]	BL00479C Phorbol esters / diacylglycerol binding domain proteins		
[BLOCKS]	BL00239B Receptor tyrosine kinase class II proteins		
[BLOCKS]	BL00107A Protein kinases ATP-binding region proteins		
[SCOP]	dlgol_5.1.1.1.9 MAP kinase Erk2 [rat Rattus norvegicus]		1e-77
[SCOP]	dlwfc_5.1.1.1.8 MAP kinase p38 [human (Homo sapiens)]		4e-68
[SCOP]	dlkoa_5.1.1.1.7 (1-350) Twitchin, kinase domain [Caenorhabditis]		2e-85
[SCOP]	dlkoba_5.1.1.1.6 Twitchin, kinase domain [california sea har]		1e-80
[SCOP]	dlphk_5.1.1.1.5 gamma-subunit of glycogen phosphorylase kinase		2e-76
[SCOP]	dlirk_5.1.1.2.4 insulin receptor [Human (Homo sapiens)]		1e-69
[SCOP]	dlapme_5.1.1.1.4 cAMP-dependent PK, catalytic subunit [mouse (Mu)]		1e-84
[SCOP]	dlfgka_5.1.1.2.3 Fibroblast growth factor receptor 1 [human (Homo)]		1e-68
[SCOP]	dlydre_5.1.1.1.3 cAMP-dependent PK, catalytic subunit [bovine (Bo)]		9e-85
[SCOP]	dlfmk_5.1.1.2.2 (168-437) c-src tyrosine kinase [human (Homo)]		1e-69
[SCOP]	dicdka_5.1.1.1.2 cAMP-dependent PK, catalytic subunit [pig (Su)]		1e-85
[SCOP]	d2hcka_5.1.1.2.1 (167-437) Haemopoietic cell kinase Hck [human]		5e-66
[SCOP]	dlcsn_5.1.1.1.11 Casein kinase-1, CK1 [Schizosaccharomyces pombe]		9e-47
[SCOP]	dljsua_5.1.1.1.1 Cyclin-dependent PK [Human (Homo sapiens)]		1e-75
[SCOP]	dlckja_5.1.1.1.10 Casein kinase-1, CK1 [rat (Rattus norvegicus)]		5e-54
[EC]	2.7.1.38 Phosphorylase kinase		1e-36
[EC]	2.7.1.123 Ca2+/calmodulin-dependent protein kinase		4e-40

[EC] 2.7.1.128 [Acetyl-CoA carboxylase] kinase 1e-61
 [EC] 2.7.1.117 Myosin-light-chain kinase 2e-40
 [EC] 2.7.1.109 [Hydroxymethylglutaryl-CoA reductase(NADPH)] kinase 1e-61
 [EC] 2.7.1.37 Protein kinase 7e-42
 [PIRKW] phosphotransferase 6e-66
 [PIRKW] nucleus 1e-64
 [PIRKW] calcium 7e-35
 [PIRKW] duplication 1e-38
 [PIRKW] tandem repeat 4e-39
 [PIRKW] phorbol ester binding 1e-38
 [PIRKW] zinc 1e-38
 [PIRKW] cell cycle control 1e-42
 [PIRKW] serine/threonine-specific protein kinase 8e-68
 [PIRKW] oncogene 1e-40
 [PIRKW] phospholipid binding 1e-38
 [PIRKW] autophosphorylation 1e-64
 [PIRKW] brain 1e-40
 [PIRKW] heterotetramer 2e-36
 [PIRKW] mitosis 7e-42
 [PIRKW] polymer 1e-35
 [PIRKW] magnesium 6e-66
 [PIRKW] ATP 8e-68
 [PIRKW] polyprotein 1e-40
 [PIRKW] phosphoprotein 1e-64
 [PIRKW] apoptosis 4e-39
 [PIRKW] glycoprotein 7e-42
 [PIRKW] leucine zipper 3e-35
 [PIRKW] skeletal muscle 7e-35
 [PIRKW] protein kinase 5e-41
 [PIRKW] cAMP binding 3e-38
 [PIRKW] testis 9e-36
 [PIRKW] purine nucleotide binding 2e-49
 [PIRKW] calcium binding 8e-39
 [PIRKW] alternative splicing 3e-37
 [PIRKW] P-loop 2e-49
 [PIRKW] lipoprotein 2e-33
 [PIRKW] segmentation 1e-33
 [PIRKW] core protein 1e-40
 [PIRKW] muscle 7e-35
 [PIRKW] myristylation 2e-33
 [PIRKW] EF hand 8e-39
 [PIRKW] cell division 2e-40
 [PIRKW] calmodulin binding 4e-40
 [SUPFAM] ribosomal protein S6 kinase II 5e-36
 [SUPFAM] fibronectin type III repeat homology 3e-33
 [SUPFAM] immunoglobulin homology 3e-33
 [SUPFAM] calcium-dependent protein kinase 8e-39
 [SUPFAM] AMP-activated protein kinase 6e-66
 [SUPFAM] protein kinase akt 3e-42
 [SUPFAM] protein kinase SPK1 1e-42
 [SUPFAM] unassigned Ser/Thr or Tyr-specific protein kinases 8e-68
 [SUPFAM] Ca2+/calmodulin-dependent protein kinase 3e-37
 [SUPFAM] calmodulin repeat homology 8e-39
 [SUPFAM] cAMP receptor protein cyclic nucleotide-binding domain homology 6e-33
 [SUPFAM] protein kinase C zeta 1e-36
 [SUPFAM] Dictyostelium cAMP-dependent protein kinase catalytic chain 1e-34
 [SUPFAM] death-associated protein kinase 4e-39
 [SUPFAM] pleckstrin repeat homology 3e-42
 [SUPFAM] ankyrin repeat homology 4e-39
 [SUPFAM] protein kinase homology 8e-68
 [SUPFAM] Ca2+/calmodulin-dependent protein kinase II 8e-41
 [SUPFAM] protein kinase C zinc-binding repeat homology 1e-38
 [SUPFAM] twitchin 3e-33
 [SUPFAM] protein kinase C delta 1e-38
 [SUPFAM] cGMP-dependent protein kinase 6e-33
 [SUPFAM] protein kinase cdr1 7e-42
 [SUPFAM] protein kinase C C2 region homology 3e-37
 [SUPFAM] protein kinase C alpha 3e-37
 [SUPFAM] yeast protein kinase C 5e-36
 [SUPFAM] kinase-related transforming protein 1e-41
 [SUPFAM] kinase interaction domain homology 1e-42
 [SUPFAM] gag-akt polyprotein 1e-40
 [SUPFAM] Ca2+/calmodulin-dependent protein kinase I 4e-40
 [SUPFAM] protein kinase C mu 4e-33
 [PROSITE] PROTEIN_KINASE_ATP 2
 [PROSITE] RGD 1
 [PROSITE] MYRISTYL 4
 [PROSITE] CAMP_PHOSPHO_SITE 3
 [PROSITE] CK2_PHOSPHO_SITE 13
 [PROSITE] TYR_PHOSPHO_SITE 2
 [PROSITE] PKC_PHOSPHO_SITE 12

```

[PROSITE]      ASN_GLYCOSYLATION      2
[PROSITE]      PROTEIN_KINASE_ST      1
[PFAM]         Eukaryotic protein kinase domain
[KW]           All_Alpha
[KW]           3D
[KW]           LOW_COMPLEXITY      10.51 %

```

```

SEQ      MESLVFARRSGPTPSAAELARPLAELIKSPKPLMKKQAVKRHHKHNLRHRYEFLETIG
SEG      .....XXXXXXXXXXXXX.....
lctpE    .....HHHHHHHHHHHHHCCCCCCCC--GGGEEEEEEEE

SEQ      KGTYGKVKKARESSGRLVAIKSIRKDKIKDEQDLMHIRREIEIMSSLNHPHIIAIHEVFE
SEG      .....
lctpE    CTTTEEEEEETTTTEEEEEEEHHHHHHHCCCHHHHHHHHHHHHCCCTTTBCCEEEEEEE

SEQ      NSSKIVIVMEYASRGDLYDYISERQQLSEREARHFFRQIVSAVHYCHQNRVVRDLKLEN
SEG      .....
lctpE    ETTEEEEEECTTTTBHHHHHHHHHCCCCHHHHHHHHHHHHHHHHHHHCCCECCCCGGG

SEQ      ILLDANGNIKIADFGLSNLYHQKFLQTFGSPLYASPEIVNGKPYTGPEVDSWSLGVLL
SEG      .....
lctpE    EEETTTTCEEECTTTTTEET-TTT-BCCCCCGGGCCHHHHHCCCBBC-HHHHHHHHHHHH

SEQ      YILVHGTMFDPGDHDKILVKQISNGAYREPPKPSDACGLIRWLLMVNPTRRATLEDVASH
SEG      .....
lctpE    HHHHHHCTTTTTHHHHHHHHHHCCCCCTTTCHHHHHHHHHHTTTTGGGTTTHHHHHHHC

SEQ      WWWNGYATRVEQEAPHEGGHSGSDSARASMADWLRRSSRPLENGAKVCSFFKQHAPG
SEG      .....
lctpE    GG.....

SEQ      GGSTTPGLERQHSCLKSRKENDMAQSLHSDTADDTAHRPGKSNLKLPGKILKKKVSASAE
SEG      .....
lctpE    .....

SEQ      GVQEDPELSPIPASPGQAAPLLPKKGILKKPRQRESGYSSPEPSESSELLDAGDVVFS
SEG      .....XXXXXXXXXXXXX.....XXXXXXXXXXXXX.....
lctpE    .....

SEQ      GDPKEQKPPQASGLLLHRKGILKLNKFSQTALELAAPTTFGSLDELAPPRPLARASRPS
SEG      .....XXXXXXXXXXXXX.....
lctpE    .....

SEQ      GAVSEDSILSSESFQDLQDLPERLPEPPLRGCVSDNLTGLEEPPSEGPGLRRWRQDPL
SEG      .....XXXXXXXXXXXXX.....
lctpE    .....

SEQ      GDSCFSLTDCQEVATYRQALRVCSKLT
SEG      .....
lctpE    .....

```

Prosites for DKFZphtes3_7j3.2

PS00001	121->125	ASN_GLYCOSYLATION	PDOC00001
PS00001	576->580	ASN_GLYCOSYLATION	PDOC00001
PS00004	290->294	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	337->341	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	413->417	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	30->33	PKC_PHOSPHO_SITE	PDOC00005
PS00005	74->77	PKC_PHOSPHO_SITE	PDOC00005
PS00005	82->85	PKC_PHOSPHO_SITE	PDOC00005
PS00005	122->125	PKC_PHOSPHO_SITE	PDOC00005
PS00005	142->145	PKC_PHOSPHO_SITE	PDOC00005
PS00005	148->151	PKC_PHOSPHO_SITE	PDOC00005
PS00005	289->292	PKC_PHOSPHO_SITE	PDOC00005
PS00005	327->330	PKC_PHOSPHO_SITE	PDOC00005
PS00005	339->342	PKC_PHOSPHO_SITE	PDOC00005
PS00005	373->376	PKC_PHOSPHO_SITE	PDOC00005
PS00005	377->380	PKC_PHOSPHO_SITE	PDOC00005
PS00005	616->619	PKC_PHOSPHO_SITE	PDOC00005
PS00006	15->19	CK2_PHOSPHO_SITE	PDOC00006
PS00006	133->137	CK2_PHOSPHO_SITE	PDOC00006
PS00006	148->152	CK2_PHOSPHO_SITE	PDOC00006
PS00006	227->231	CK2_PHOSPHO_SITE	PDOC00006
PS00006	293->297	CK2_PHOSPHO_SITE	PDOC00006
PS00006	331->335	CK2_PHOSPHO_SITE	PDOC00006
PS00006	377->381	CK2_PHOSPHO_SITE	PDOC00006
PS00006	391->395	CK2_PHOSPHO_SITE	PDOC00006

PS00006	461->465	CK2_PHOSPHO_SITE	PDOC00006
PS00006	511->515	CK2_PHOSPHO_SITE	PDOC00006
PS00006	523->527	CK2_PHOSPHO_SITE	PDOC00006
PS00006	578->582	CK2_PHOSPHO_SITE	PDOC00006
PS00006	606->610	CK2_PHOSPHO_SITE	PDOC00006
PS00007	453->460	TYR_PHOSPHO_SITE	PDOC00007
PS00007	453->461	TYR_PHOSPHO_SITE	PDOC00007
PS00008	320->326	MYRISTYL	PDOC00008
PS00008	324->330	MYRISTYL	PDOC00008
PS00008	347->353	MYRISTYL	PDOC00008
PS00008	360->366	MYRISTYL	PDOC00008
PS00016	134->137	RGD	PDOC00016
PS00107	59->82	PROTEIN_KINASE_ATP	PDOC00100
PS00107	59->86	PROTEIN_KINASE_ATP	PDOC00100
PS00108	171->184	PROTEIN_KINASE_ST	PDOC00100

Pfam for DKFZphtes3_7j3.2

HMM_NAME	Eukaryotic protein kinase domain		
HMM	*YeigRiIGeGsFGtVYkCiWrTGeIVAiKIIkkrsms.....FlREI		
	YE+++++G+G+G+V+K+++ +G++VAIK I+K++++ ++REI		
Query	53	YEFLETLGKGTGKVKKARESSGRLVAIKSIRKDKIKDEQDLMHIRREI	101
HMM	qIMRrLnHPNIIRFYDwFedddDHIYMIMEYMeGGDLFDYIrrngpmSEw		
	+IM +LNHP+II + ++FE ++ I ++MEY+ GDL+DYI+++ ++SE+		
Query	102	EIMSSLNHPHIIAIHEVFE-NSSKIVIVMEYASRGDLYDYISERQQLSER	150
HMM	eIrfIMyQILrGMeYLHSMgIIHRDLKPENILIDeNgqIKicDFGLARqM		
	E+R++++QI++++ Y+H ++++HRDLK ENIL+D NG+IKI+DFGL+ ++		
Query	151	EARHFFRQIVSAVHYCHQNRVVRDLKLENILLDANGNIKIADFGLSNLY	200
HMM	nnYerMttfCGTPWYMAPEVIImg.nyYttkVDMWSFGCILWEMMTGep		
	+ + ++ TFCG+P Y +PE+ ++G +Y +++VD WS+G++L++++ G+		
Query	201	HQGKFLQTFCGSPLYA-SPEI-VNGKPYTGPEVDSWSLGVLLYILVHGTM	248
HMM	PFyddnMemImrIiqrfrrpfWpncSeElyDFMrWCWnyDPekRPTFrQI		
	PF+++ ++ I + +++ +P S+ + ++RW++ ++P++R T +++		
Query	249	PFdGHDHKILVKQISNGAYREPPKPSD-ACGLIRWLLMVNPTRRATLEDV	297
HMM	LnHPWF*		
	H W+		
Query	298	ASHWWV 303	

DKFZphtes3_7j8

group: testes derived

DKFZphtes3_7j8 encodes a novel 410 amino acid protein nearly identical to human
WUGSC:H_DJ1159004.1.

The novel protein contains an additional C-terminal domain, which is not present in
WUGSC:H_DJ1159004.1.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific
genes.

WUGSC:H_DJ1159004.1 similarity to YBL104p

verifies and extends the genmodel WUGSC:H_DJ1159004.1
similarity to *S.cerevisiae* YBL104p

Sequenced by BMFZ

Locus: /map="7p21-p22"

Insert length: 3353 bp

Poly A stretch at pos. 3231, no polyadenylation signal found

```

1  GCAAAATATG TTGTATTTGT GGCATAGTTC ATATTTACAC TATCATAAAA
51  TTATGGCCGA GAAGTTAAAT ATTCTAAATG TGTCACATA GTTCTCTGTA
101 AAACCTGACTT ATTTTCCAAA TATATTTTGA AATAAAACAA TATAAAAATG
151 TTTTCTGTTT TTAGGAATGG TGGAAAGCAG CAGACATAAT TGGAGTGGGT
201 TGGATAAGCA AAGTGATATT CAAAATTTAA ATGAAGAGAG AATCTTAGCT
251 TTACAGCTTT GTGGGTGGAT AAAGAAAGGA ACGGATGTAG ACGTGGGGCC
301 ATTTTGAAC TCCTTTGTAC AAGAAGGGGA ATGGGAAAGA GCTGCTGCTG
351 TGGCATTGTT CAACCTGGAT ATTCGCCGAG CAATCCAAAT CCTGAATGAA
401 GGGGCATCTT CTGAAAAAGG AGATCTGAAT CTCAATGTGG TAGCAATGGC
451 TTTATCGGGT TATACGGATG AGAAGAACTC CCTTTGGAGA GAAATGTGTA
501 GCACACTGCG ATTACAGCTA AATAACCCGT ATTTGTGTGT CATGTTTGCA
551 TTTCTGACAA GTGAACAGG ATCTTACGAT GGAGTTTGT ATGAAAACAA
601 AGTTGCAGTA CGTGACAGAG TGGCATTGTC TTGTAAATTC CTTAGTGATA
651 CTCAGTTAAA TAGATACATC GAAAAGTTGA CCAATGAAAT GAAAGAGGCT
701 GGAATTTTGG AAGGAATTTT GCTTACAGGC CTTACTAAAG ATGGAGTGGG
751 CTTAATGGAG AGTTATGTTG ATAGAAGTGG AGATGTTCAA ACAGCAAGTT
801 ACTGTATGTT ACAGGGTTCA CCTTTAGATG TTCTTAAAGA TGAAAGGGTT
851 CAGTACTGGA TTGAGAATTA TAGAAATTTA TTAGATGCCT GGAGGTTTTG
901 GCATAAACGA GCTGAATTTG ATATTCACAG GAGTAAGTTG GATCCCAGTT
951 CCAAGCCTTT AGCACAAAGT TTTGTGAGTT GCAATTTCTG TGGCAAGTCA
1001 ATCTCCTACA GCTGTTTCAGC TGTGCCTCAT CAGGGCAGAG GTTTTAGTCA
1051 GTTGTGTTG AGTGGCTCAC CAACGAAATC TAAAGTCACA AGTTGTCCTG
1101 GCTGTGCGAA ACCACTTCCT CGATGTGCGC TTTGTCTCAT TAATATGGGA
1151 ACACCAAGTT CTAGCTGTCC TGGAGGAACC AAATCAGATG AAAAAGTGGA
1201 CTTGAGCAAG GACAAAAAAT TAGCCCAATT TAACAACCTG TTTACATGGT
1251 GTCATAATTG CAGGCACGGT GGACATGCTG GACATATGCT TAGTTGGTTC
1301 AGGGACCATG CAGAGTGCCC TGTGTCTGCA TGCACGTGTA AATGTATGCA
1351 GTTGGATACA ACGGGGAATC TGGTACCTGC AGAGACTGTC CAGCCATAAA
1401 ATGTTACCAC CTTAAGAGAA CCCTTCAAGT GTGGAGCTTT CTAGTAGGTG
1451 TCCTTCATAG CTCAGAAACA TACCTCAGAA CAAGCCATTC ATGACTTACC
1501 TGTAATGGGA AAATAAATCA TTCTATCAGA TCAGCAGTTT TGATGTTTGA
1551 GTGATTTTGA TATGCTTCAC AGAGACAAAT GCTGCCAAAA TAAACATCGA
1601 AGTATAGACA TGAGTTCTGT TCAGCAGGTT GAAAAGTCTG ATTTAGAAAA
1651 ACTTTCTAAG TTTTGGTTGA AATTATGAAC ACTCTAGAAG CAGAATTTCT
1701 GGAAGAGCCA AGAACAGACT TTGAGCCTAT ATCTTCAAAG CTGAAACTGG
1751 ATATCTTTCA ATAAAATATG TGCACTTTAA AAATAAAATG ACTAATTCTG
1801 TGATTCAGAC AATAGTTTTA AGTTCAGCTG TGCTTAGATT TCTTTCAGAT
1851 TAATTTAAAA TTATAGATTT TTAATTTTAG AATTCAGAG CCCCTATCCC
1901 AACTGGAGA ATATTTTTTA TTACTGTCTG TTATATATGT GTCTATGTGT
1951 GTGTGTATAT TTATGTGTGT ATGTATAAAT ATGTACTTTT TAAAGGAGCC
2001 TTTTCCCTCC TTTGATTTTA AGATAAGCAA TCTTTTGGCA TAACATTATC
2051 GTCTTCCTAG AAAAGCCAAG ATGAAGAATC TATCTTACAA CTTTTCCTCT
2101 TCAGTAGAGA AAAACATGTA CCATTTCAAG TGAACATACA AAATTTTCAC
2151 TTTCTACCTT TTGCCTTCCA ATGTCCTGAT TTGCTTCAA AGGTTTTTCT
2201 CCATATTAAT TTGTCATCTT ATCCTCATCA CCTGAGAACA TTTTACTGCA
2251 TACAAAGTCT ATGCAAGATT ATATGTAAC AGCCATTTAG TATAATCTAT
2301 GTCAGTGTGT CTGTGCTGTC AAATCCGTC CTGATTTGGA ATACCATACC
2351 TTGTTCTTTC CAAGGTAGAC TAGGAAAGTG TGGGGAATA GGGTCACTTC
2401 AGAGACCATT TTAGATGTAA GTTTTTAAAT GTAAGTGTTA CTGGGGCTAA
2451 GTCAGGACTT TTATTTAAAA CATTTTTTTT TTCTCATTTT ATAGCTAGAT
2501 AGTTGTAAGA GAAATACAAA GAATTTACAA GATGCTTCTC TGTCATCTGC

```

```

2551 CGTATGCAGA GGGACTGAAC TAGGAATTTT GTAGTTGAAG CTGTGTTTCAT
2601 AAAGAGTAAA TCTTATTTTA TAGATTTTGG AGAAATAAAA CAAGAATTTT
2651 AAGAGCTTTC GTATTAGCAG TTTTGCCCTTA TAAAAACTAA GATTTGTCAG
2701 ATTAGTTTGA GGTGTAACCT AAATATTTAA AGTAGATTAA ATTTATTTT
2751 TACCTTGAGT GTCTGATACA TAAAACCCCT TTCTAGGAAA ACATTGGAAG
2801 TAGTACATAT TACTCTAAA TGTCTCACCT GCATGACAGT CTTTCAAAT
2851 GAAAGACATG GTAATTGCAA TTTTTTTTA AAGATTGCTA TTAAGGGTAC
2901 TTTTCCAGC CTTCATTGTA GTAAATCTTA ATTGATTTC TTTTATTAAC
2951 ATATACCCCT TACCTTTAAT ATTTCAATTG AAGTGTTCTC TTCAAACTTA
3001 CTGTCTTAAA TATGAAAGTC AGCTTTAAGT AATGTCAGAC TCATATGCAT
3051 TTTCACTCTC ATTAGCTAAA GTAAATGTGA AAATTATCTC AAATAGTTAC
3101 AAGTTTGGGA AATACAGTAT AAAACATGAA TGTAAAGTCT ATTATGTAAT
3151 ATGCTTATTT GTAATCCTAA TATATGAGGG TGACATTTT AAGATTGTAT
3201 GTATGTGTCA ACCTCTTAAA TGTTTCTGT GAAAAAAAAA AAAAAAAAAA
3251 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3301 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3351 AAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 167 bp to 1396 bp; peptide length: 410
 Category: known protein
 Classification: unclassified

```

1 MVESSRHNS GLDKQSDIQN LNEERILALQ LCGWIKKGT DVGPFNLNSL
51 VQEGEWERAA AVALFNLDIR RAIQILNEGA SSEKGDNLN VVAMALSGYT
101 DEKNSLWREM CSTLRLQLNN PYLCVMFAFL TSETGSYDGV LYENKVAVRD
151 RVAFAACKFLS DTQLNRYIEK LTNEMKEAGN LEGILLTGLT KQGVDLMESY
201 VDRTGDVQTA SYCMLQGSPL DVLKDERVQY WIENYRNLLD AWRFWHKRAE
251 FDIHRSKLDP SSKPLAQVFV SCNFCGKXIS YSCSAVPHQ RGFSQYGVSG
301 SPTKSKVTSC PGCRLPLPRC ALCLINMGTP VSSCPGGTKS DEKVDLSKDK
351 KLAQFNWFT WCHNCRHGGH AGHMLSWFRD HAECVPVSACT CKCMQLDITG
401 NLVPAETVQP

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_7j8, frame 2

PIR:S45391 probable membrane protein YBL104c - yeast (*Saccharomyces cerevisiae*), N = 2, Score = 446, P = 4.5e-47

TREMBL:AC004982_1 gene: "WUGSC:H_DJ1159004.1"; Homo sapiens PAC clone
 DJ1159004 from 7p21-p22, complete sequence., N = 1, Score = 2038, P = 7.6e-211

>TREMBL:AC004982_1 gene: "WUGSC:H_DJ1159004.1"; Homo sapiens PAC clone
 DJ1159004 from 7p21-p22, complete sequence.
 Length = 379

HSPs:

Score = 2038 (305.8 bits), Expect = 7.6e-211, P = 7.6e-211
 Identities = 379/379 (100%), Positives = 379/379 (100%)

```

Query:      1 MVESSRHNSGLDKQSDIQNLNEERILALQLCGWIKKGTVDVGPFLNSLVQEGEWERAA 60
            MVESSRHNSGLDKQSDIQNLNEERILALQLCGWIKKGTVDVGPFLNSLVQEGEWERAA
Sbjct:      1 MVESSRHNSGLDKQSDIQNLNEERILALQLCGWIKKGTVDVGPFLNSLVQEGEWERAA 60

Query:     61 AVALFNLDIRRAIQILNEGASSEKGDNLNVVAMALSGYTDEKNSLWREMCSTLRLQLNN 120
            AVALFNLDIRRAIQILNEGASSEKGDNLNVVAMALSGYTDEKNSLWREMCSTLRLQLNN
Sbjct:     61 AVALFNLDIRRAIQILNEGASSEKGDNLNVVAMALSGYTDEKNSLWREMCSTLRLQLNN 120

```

Query: 121 PYLCVMFAFLTSETGSYDGVLYENKVAVRDRVAFACKFLSDTQLNRYIEKLTNEMKEAGN 180
 PYLCVMFAFLTSETGSYDGVLYENKVAVRDRVAFACKFLSDTQLNRYIEKLTNEMKEAGN
 Sbjct: 121 PYLCVMFAFLTSETGSYDGVLYENKVAVRDRVAFACKFLSDTQLNRYIEKLTNEMKEAGN 180

Query: 181 LEGILLTGLTKDGVLDLMESYVDRTGDVQTASYCMLQGSPLDVLKDERVQYWIENYRNLLD 240
 LEGILLTGLTKDGVLDLMESYVDRTGDVQTASYCMLQGSPLDVLKDERVQYWIENYRNLLD
 Sbjct: 181 LEGILLTGLTKDGVLDLMESYVDRTGDVQTASYCMLQGSPLDVLKDERVQYWIENYRNLLD 240

Query: 241 AWRFWHKRAEFDIHRSKLDPSSKPLAQVFVSCNFCGKSISYSCSAVPHQGRGFSQYGVSG 300
 AWRFWHKRAEFDIHRSKLDPSSKPLAQVFVSCNFCGKSISYSCSAVPHQGRGFSQYGVSG
 Sbjct: 241 AWRFWHKRAEFDIHRSKLDPSSKPLAQVFVSCNFCGKSISYSCSAVPHQGRGFSQYGVSG 300

Query: 301 SPTKSKVTSCPGCRKPLPRCALCLINMGTPVSSCPGGTKSDEKVDLSKDKKLAQFNNWFT 360
 SPTKSKVTSCPGCRKPLPRCALCLINMGTPVSSCPGGTKSDEKVDLSKDKKLAQFNNWFT
 Sbjct: 301 SPTKSKVTSCPGCRKPLPRCALCLINMGTPVSSCPGGTKSDEKVDLSKDKKLAQFNNWFT 360

Query: 361 WCHNCRHGGHAGHMLSWFR 379
 WCHNCRHGGHAGHMLSWFR
 Sbjct: 361 WCHNCRHGGHAGHMLSWFR 379

Pedant information for DKFZphtes3_7j8, frame 2

Report for DKFZphtes3_7j8.2

[LENGTH] 410
 [MW] 45862.45
 [pI] 6.51
 [HOMOL] TREMBL:AC004982_1 gene: "WUGSC:H_DJ1159004.1"; Homo sapiens PAC clone DJ1159004
 from 7p21-p22, complete sequence. 0.0
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YBL104c] 7e-48
 [BLOCKS] BL00028 Zinc finger, C2H2 type, domain proteins
 [BLOCKS] BL00534A Ferrochelatae proteins
 [PIRKW] transmembrane protein 2e-46
 [KW] All_Alpha

SEQ MVESRRHNWGLDKQSDIQNLNEERILALQLCGWIKKGTVDVGFPLNSLVQEGEWERA
 PRD cccccccccccccccccchhhhhhhhhhhccccccccccccccccccccccccchhhh

SEQ AVALFNLDIRRAIQILNEGASSEKGDNLNVVAMALSGYTDEKNSLWREMCSTLRLQLNN
 PRD hhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhcc

SEQ PYLCVMFAFLTSETGSYDGVLYENKVAVRDRVAFACKFLSDTQLNRYIEKLTNEMKEAGN
 PRD cccceeeccccccccccccceccccchhhhhhhhhhhhhccccchhhhhhhhhhhhhhhcc

SEQ LEGILLTGLTKDGVLDLMESYVDRTGDVQTASYCMLQGSPLDVLKDERVQYWIENYRNLLD
 PRD cceeeeeeccccchhhhhhhhhccccceeeccccccccccccchhhhhhhhhhhhhhhhh

SEQ AWRFWHKRAEFDIHRSKLDPSSKPLAQVFVSCNFCGKSISYSCSAVPHQGRGFSQYGVSG
 PRD hhhhhhhhhhhhhhhccccccccceeecccccccccccccccccccccccccccccccc

SEQ SPTKSKVTSCPGCRKPLPRCALCLINMGTPVSSCPGGTKSDEKVDLSKDKKLAQFNNWFT
 PRD cccccccccccccccccceeeccccccccccccccccccccceeehhhhhhhhhhccccc

SEQ WCHNCRHGGHAGHMLSWFRDHAECPVSACTCKCMQLDTTGNLVPAAETVQP
 PRD eccccccccccccchhhhhhhhhcccccccccccccccccccccccccccccccccccc

(No Prosite data available for DKFZphtes3_7j8.2)

(No Pfam data available for DKFZphtes3_7j8.2)

DKFZphtes3_7p10

group: Cell Cycle

DKFZphtes3_7p10.1 encodes a novel 422 amino acid putative protein, which is closely related to the *Xenopus laevis* XPMC2 protein.

In fission yeast the kinases Wee1 and Mik1 control that initiation of mitosis starts after completion of DNA synthesis. Yeast in which both Wee1 and Mik1 kinases are defective exhibit a mitotic catastrophe phenotype. XPMC2 of *xenopus* rescues several different yeast mitotic catastrophe mutants defective in Wee1/Mik1 kinase function. The XPMC2 protein is localised in the nucleus in *Xenopus* oocytes. The new protein is the human orthologue of this gene.

The new protein can find application in modulating/blocking the cell cycle.

strong similarity to XPMC2 protein

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: /map="9q34"

Insert length: 2380 bp

Poly A stretch at pos. 2341, polyadenylation signal at pos. 2318

```

1  AGCGTGCCTG  CTGAGGTATG  CGCAACGCGT  GCGGGGTCTC  TTCCGGAGTC
51  TTTTCCTGGA  CCGGGTCCCT  GCGGTGGGTG  TGTTCGGGCC  TGGCCTGGGC
101 AGGCGCTTGT  GCTGCCAGGG  CGCCGGGCCC  GGGGAGGCCG  GGGTCTCGGG
151 TGGCCGCCGG  CCCAGGCGCT  GGACGGCAGC  AGGATGGGGA  AGGCGAAGGT
201 CCCCCTCTCC  AAGCGCGCCC  CGAGCAGCCC  CGTGGCTAAG  CCGGGTCTTG
251 TCAAGACGCT  CACTCGGAAG  AAAACAAGA  AGAAAAAAG  GTTTTGGA
301 AGCAAGGCGC  GGGAAATAAG  CAAGAAGCCA  GCAAGCGGCC  CCGGTGCTGT
351 GGTGCGACCT  CCAAAGGCAC  CAGAAGACTT  TTCTCAAAAC  TGGAAAGCGC
401 TGCAAGAGTG  GCTGCTGAAA  CAAAAATCTC  AGGCCCCAGA  AAAGCCTCTT
451 GTCATCTCTC  AGATGGGTTC  CAAAAAGAAG  CCCAAATTA  TCCAGCAAAA
501 CAAAAAAGAG  ACCTCGCCTC  AAGTGAAGGG  AGAGGAGATG  CCGGCAGGAA
551 AAGACCAGGA  GGCCAGCAGG  GGCTCTGTTC  CTTCAGGTTC  CAAGATGGAC
601 AGGAGGGCGC  CAGTACCTCG  CACCAAGGCC  AGTGAACAG  AGCACAATAA
651 GAAAGGAACC  AAGGAAGGA  CAAATGGTGA  TATTGTTCCA  GAACGAGGGG
701 ACATCGAGCA  TAAGAAGCGG  AAAGCTAAGG  AGGCAGCCCC  AGCCCCACCC
751 ACCGAGGAAG  ACATCTGGTT  TGACGACGTG  GACCCAGCGG  ATATCGAAGC
801 TGCCATAGGT  CCAGAGGCGG  CCAAGATAGC  GAGGAAACAG  TTGGGTGAGA
851 GCAGGGGCGC  CGTCAGCCTC  AGCCTCGTGA  AAGAGCAGGC  CTTGCGCGGC
901 CTGACAAGAG  CCTAGCCTT  GGACTGTGAG  ATGGTGGGCG  TGGGCCCTAA
951 GGGGGAGGAG  AGCATGGCCG  CCGTGTGTC  CATCGTGAAC  CAGTATGGGA
1001 AGTGCCTTTA  TGACAAGTAC  GTCAAAACCA  CTGAGCCCGT  GACGGACTAT
1051 AGGACAGCGG  TCAGTGGGAT  TCGGCCTGAG  AACCTCAAGC  AGGGAGAAGA
1101 GCTTGAAGTT  GTTCAGAAGG  AAGTGGCAGA  GATGCTGAAG  GGCAGAATTC
1151 TAGTGGGGCA  CGCTCTGCAT  AATGACCTAA  AGGTACTATT  TCTTGATCAT
1201 CCAAAAAAGA  AGATTGCGGA  CACACAGAAA  TATAAACCTT  TCAAGAGTCA
1251 AGTAAAGAGT  GGAAGGCCGT  CTCTGAGACT  ACTTTCAGAG  AAGATCCTTG
1301 GGCTCCAGGT  CCAGCAGGCG  GAGCACTGTT  CAATTCAGGA  TGCCAGGCA
1351 GCAATGAGGC  TGTACGTCAT  GGTGAAGAAG  GAGTGGGAGA  GCATGGCCCG
1401 AGACAGGCGC  CCCCTGCTGA  CTGCTCCAGA  CACTGCAAGT  GACGACGCTT
1451 AGCAGTCCCT  CCCTGCTGCT  GCTGCCGCCC  CGCTACAGAG  GCAATGTGAC
1501 CAGTCACAGG  GACAGATCAC  ATCTCCCCAG  AGTGGCAACT  CTGGTGAAC
1551 CTTTTCAGAA  TCATGGCAGA  GGGCGTGGC  GTGGTGTAC  TGAGAAGGTC
1601 CTCCTTCCTC  TTGACTTTGT  GGTCTGAAAC  CTGGTCTTAC  TGTCCATGTG
1651 TGTTTGGGCC  CGGATGGTCA  GGGTGGGGAG  CAGGGACGGC  CATGGGCACG
1701 CTTGGCCACG  CTTTACCGAC  TGCTGACCCC  CTGGGCCAGG  TGAGGTTGGG
1751 GCCTGTGGGC  CGCCAGTCCA  TACGGTGCTG  TCACTGCCCA  TCTTCGGTGA
1801 CACCTGGGG  TGAGGTGCTC  AGCACCTTCC  TCTCGAGGAG  CCACATTTTC
1851 CTCCTTTGTG  TTAGGGGACA  TAACAAGCTC  TGCTGGGCTT  GAGGGACCCA
1901 GACCAGGTGT  CTGCAGTCAG  CTCCTGAGAC  ACAGCTGGCC  GGCACACAG
1951 GTGTTACATC  AGGGGTTTCC  TGTGGCCGTT  TGAACCTTGA  GCATTTATCT
2001 AAATTAATTT  GGCCAGGGT  TGGCTGTGTG  GTCACCCAGC  AGAGGCTTCT
2051 CCCCATAGCA  CGAGGATGTG  TTGCCCTGGG  ACGGTGACTG  CGGTATTCC
2101 TGGAGGTCGG  CAGACATGCC  AACCTTGGGC  TATTTGAGCT  GGAGAAGCTA
2151 TGTGATGCTA  GCCGTTGGCT  TTCTGGGCTA  GGCCCAAGTT  TGAGGCTCCC
2201 CTGGGAACCT  GAGCCAGGAA  CAGCCAGTGG  CACTGACAAG  GGGACGGAGT
2251 CCAAGGCGTT  ATTGGGCCAC  CTGACAGCTG  GACAGAAAAG  GGGCAGACAC
2301 ACCGAGGATG  CGATTTAAAA  TAAATGCAGA  TGTTTACTTG  GAAAAA
2351 AAAAAA

```

BLAST Results

Medline entries

Peptide information for frame 1

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3 7p10, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3 7p10, frame 1

Report for DKFZphtes3 7p10.1

```

[LENGTH]      422
[MW]           46671.91
[pI]           9.79
[HOMOL]        PIR:S53818 XPMC2 protein - African clawed frog 7e-96
[FUNCAT]       03.22 cell cycle control and mitosis [S. cerevisiae, YOL080c] 2e-42
[FUNCAT]       01.03.16 polynucleotide degradation [S. cerevisiae, YGR276c] 2e-19
[FUNCAT]       05.04 translation (initiation, elongation and termination) [S. cerevisiae,
YGL094c] 7e-13
[FUNCAT]       04.05.05 mrna processing (5'-end, 3'-end processing and mrna degradation) [S.
cerevisiae, YGL094c] 7e-13
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YLR107w] 6e-10
[PROSITE]      RGD 1
[PROSITE]      MYRISTYL 4
[PROSITE]      CAMP_PHOSPHO_SITE 2
[PROSITE]      CK2_PHOSPHO_SITE 6
[PROSITE]      TYR_PHOSPHO_SITE 2
[PROSITE]      GLYCOSAMINOGLYCAN 1
[PROSITE]      PKC_PHOSPHO_SITE 8
[KW]           All Alpha
[KW]           LOW COMPLEXITY 11.37 %

```

952

```

SEQ    AGKDQEASRGSVPSGSKMDRRAPVPRTKASGTEHNKKGTKERTNGDIVPERGDIEHKRRK
SEG    .....xxxxxx
PRD    eccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhh

SEQ    AKEAAPAPPTTEEDIWFDDVDVPADIEAAIGPEAAKIARKQLGQSEGSVSLSLVKEQAFGGL
SEG    xxxxxxxxxxxx.....
PRD    hhhhhccccccccccccccccchhhhhccchhhhhhhhhhhccccchhhhhhhhhhhhhhh

SEQ    TRALALDCEMVGVGPKGEESMAARVSIVNQYGKCVYDKYVKPTEPVTDYRTAVSGIRPEN
SEG    .....
PRD    hhhccccccccccccchhhhhhhhhcccccccccccccccccccccccccccccccccccc

SEQ    LKQGEELVQKEVAEMLKGRILVGHALHNDLKVFLDHPKKKIRDTQKYKPKSQVKSG
SEG    .....
PRD    cccccchhhhhhhhhhhhhccceeeccchhhhhhhhhcccccccccccccccccccccccc

SEQ    RPSRLRLSEKILGLVQQAHECSIQDAQAAMRLYVMVKKEWESMARDRRPLLTAPDHCS
SEG    .....
PRD    chhhhhhhhhhhhhhhccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccc

SEQ    DA
SEG    ..
PRD    CC

```

Prosite for DKFZphtes3_7p10.1

PS00002	51->55	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	107->111	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	156->160	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	9->12	PKC_PHOSPHO_SITE	PDOC00005
PS00005	27->30	PKC_PHOSPHO_SITE	PDOC00005
PS00005	46->49	PKC_PHOSPHO_SITE	PDOC00005
PS00005	96->99	PKC_PHOSPHO_SITE	PDOC00005
PS00005	347->350	PKC_PHOSPHO_SITE	PDOC00005
PS00005	359->362	PKC_PHOSPHO_SITE	PDOC00005
PS00005	363->366	PKC_PHOSPHO_SITE	PDOC00005
PS00005	368->371	PKC_PHOSPHO_SITE	PDOC00005
PS00006	136->140	CK2_PHOSPHO_SITE	PDOC00006
PS00006	150->154	CK2_PHOSPHO_SITE	PDOC00006
PS00006	163->167	CK2_PHOSPHO_SITE	PDOC00006
PS00006	190->194	CK2_PHOSPHO_SITE	PDOC00006
PS00006	383->387	CK2_PHOSPHO_SITE	PDOC00006
PS00006	413->417	CK2_PHOSPHO_SITE	PDOC00006
PS00007	343->351	TYR_PHOSPHO_SITE	PDOC00007
PS00007	342->351	TYR_PHOSPHO_SITE	PDOC00007
PS00008	130->136	MYRISTYL	PDOC00008
PS00008	151->157	MYRISTYL	PDOC00008
PS00008	221->227	MYRISTYL	PDOC00008
PS00008	239->245	MYRISTYL	PDOC00008
PS00016	171->174	RGD	PDOC00016

(No Pfam data available for DKFZphtes3_7p10.1)

DKFZphtes3_7p9

group: nucleic acid management

DKFZphtes3_7p9 encodes a novel 691 amino acid protein with similarity to human nuclear domain 10 protein NDP52.

The nuclear domain (ND)10 also described as POD or Kr bodies is involved in the development of acute promyelocytic leukemia and virus-host interactions. The NDP52 protein is part of this complex structure. In vivo, NDP52 is transcribed in all human tissues, but is redistributed upon viral infection and interferon treatment. ND10 plays an important role in the viral life cycle.

The novel protein is similar to NDP52. It contains three leucine zippers and a RGD cell attachment site. This protein seems to be a novel part of the ND819) complex.

The new protein can find application in modulation of viral infections and tumour events.

similarity to nuclear domain 10 protein NDP52

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: /map="329.1 cR from top of Chr12 linkage group"

Insert length: 3003 bp

Poly A stretch at pos. 2957, no polyadenylation signal found

```

1 AAGGTGAGGG GAACAGCTGA TCCGTCTGTT GGGAGGACAG ATATCTCAAG
51 GCCAGGATGG AAGAATCACC ACTAAGCCGG GCACCATCCC GTGGTGGAGT
101 CAACTTTCTC AATGTAGCCC GGACCTACAT CCCCACACACC AAGGTGGAAT
151 GTCACTACAC CCTTCCCCCA GGCACCATGC CCAGTGCCAG TGACTGGATT
201 GGCATCTTCA AGGTGGAGGC TGCCTGTGTT CGGGATTACC ACACATTTGT
251 GTGGTCTTCC GTGCCTGAAA GTACAACTGA TGGTTCCCCC ATTCACACCA
301 GTGTCCAGTT CCAAGCCAGC TACCTGCCCC AACCAGGAGC TCAGCTCTAC
351 CAGTTCGGAT ATGTGAACCG CCAGGGCCAG GTGTGTGGGC AGAGCCCCCC
401 TTTCCAGTTC CGAGAGCCAA GGCCCATGGA TGAAGTGGT ACCCTGGAGG
451 AGGCTGATGG GGGCTCTGAC ATCCTGCTGG TTGTCCCCAA GGCAACTGTG
501 TTACAGAAC AGCTCGATGA GAGCCAGCAA GAACGGAATG ACCTGATGCA
551 GCTGAAGCTA CAGCTGGAGG GACAGGTGAC AGAGCTGAGG AGCCGAGTGC
601 AGGAGCTCGA GAGGGCTCTG GCAACTGCCA GGCAGGAGCA CACGGAGCTG
651 ATGGAACAGT ACAAGGGGAT TTCCCGGTCC CATGGGGAGA TCACAGAAGA
701 GAGGGACATC CTGAGCCGGC AACAGGGAGA CCATGTGGCA CGCATCCTGG
751 AGCTAGAGGA TGACATCCAG ACCATCAGTG AGAAAGTGCT GACGAAGGAA
801 GTGGAGCTGG ACAGGCTTAG AGACACAGTG AAGGCCCTGA CTCGGGAACA
851 AGAGAAGCTC CTTGGGCAAC TGAAAGAAGT ACAAGCAGAC AAGGAGCAAA
901 GTGAGGCTGA GCTCCAAGTG GCACAACAGG AGAACCATCA CTTAAATTTG
951 GACCTGAAGG AGGCGAAGAG CTGGCAAGAG GAGCAGAGTG CTCAGGCTCA
1001 GCGACTGAAA GACAAGGTGG CCCAGATGAA GGACACCCTA GGCCAGGCC
1051 AGCAGCGGGT GGCCGAGCTG GAGCCCTTGA AGGAGCAGCT TCGAGGGGCC
1101 CAGGAGCTTG CAGCCTCAAG CCAGCAGAAA GCCACCCTTC TTGGGAGGA
1151 GTTGGCCAGC GCAGCAGCAG CCAGGGACCG CACCATAGCC GAACTACACC
1201 GCAGCCGCCT GGAAGTGGCT GAAGTTAAGC GCAGGCTGGC TGAGCTCGGT
1251 TTGCACTTGA AGGAAGAAAA ATGCCAATGG AGCAAGGAGC GGGCAGGGCT
1301 GCTGCAGAGT GTGGAGGCAG AGAAGGACAA GATCCTGAAG CTGAGTGCAG
1351 AGATACTTCG ATTGGAGAAG GCAGTTCAGG AGGAGAGGAC CCAAAACCAA
1401 GTGTTCAAGA CTGAGCTGGC CCGGAGAAAG GATTCTAGCC TGGTACAGTT
1451 GTCAGAAAGT AAGCGGGAGC TGACAGAGCT GCGGTCAGCC CTGCGTGTGC
1501 TCCAGAAGGA AAAGGAGCAG TTACAGGAGG AGAAACAGGA ATTGCTAGAG
1551 TACATGAGAA AGCTAGAGGC CCGCCTGGAG AAGGTGGCAG ATGAGAAGTG
1601 GAATGAGGAT GCCACCACAG AGGATGAGGA GGCCGCTGTG GGGCTGAGCT
1651 GCCCGGCAGC TGTGACAGAC TCAGAGGACG AGTCCCCAGA AGACATGAGG
1701 CTCCCACCCT ATGGCCTTTG TGAGCGTGGA GACCCAGGCT CCTCTCCTGC
1751 TGGGCCTCGA GAGGCTTCTC CCCTTGTGTG CATCAGCCAG CCGGCTCCCA
1801 TTTCTCTCTA CCTCTCTGGG CCAGCTGAGG ACAGTAGCTC TGACTCGGAG
1851 GCTGAAGATG AGAAGTCAGT CCTGATGGCA GCTGTGCAGA GTGGGGGTGA
1901 GGAGGCCAAC TTAAGTCTTC CTGAAGTGGG CAGTGCCTTC TATGACATGG
1951 CCAGTGGCTT TACAGTGGGT ACCCTGTCAG AAACCAGCAC TGGGGGCCCT
2001 GCCACCCCCA CATGGAAGGA GTGTCTATC TGAAGGAGC GCTTCTCTGC
2051 TGAGAGTGAC AAGGATGCCC TGGAGGACCA CATGGATGGA CACTTCTTTT
2101 TCAGCACCCA GGACCCCTTC ACCTTTGAGT GATCTTACTC CCTGTACAT
2151 GCACAAATAC AACTCATGTC ACACACACAC TCACACACAT GCATACACTT
2201 AGGTTTCATG CCCATTTTCT ATCACACTGG GCTCCATGAT ATTCTGTTC
2251 CTAAGAACTG CTTCTGTGTG CCCTGTTTTC ATCCCAAGAT TTCTCACTTC
2301 ATCCTCTCCT ACCTGGCTCT TTTGTCCAG GAGGGGTCC TGTTCGGAAG
2351 CAGTGGCTGA ATTTATCCCC TGAAAGTGGT TTTGGAGGAA CCGGATGGA
2401 GGAGGCCCTC CCCTGTGGGA ATAGAATCGT CCACTCCTAG CCCTGGTTGC

```

```

2451 TTCTGATACA CAGCCACTGC ACACACACAC TCACACTCAC ACTCCCTTGT
2501 CTGATGCCCC AAAGCCAATT CCTGGGGCAC CCTACCTCT CTTATTGGGA
2551 GTTTCGGTTG GTTTACCTGA GTTTTCTCTG GGGTCTGCAC AGAGGCAGCA
2601 GCATGGACAT CATGGCCTCT CAGGTCCCTT TTGGTTCTCA GTTTCATTGG
2651 TTCCTCTTTC TGTTCCTCCA TTGACTTCTG TGCCCCACCC TAGCCTTTTC
2701 CATAACCTTA GGTATTCAGT TTGGAGGGGT TTTTGTATT TTTGAGGATT
2751 CCTGTATTCT GTATCCTCTC CTCGCATCTC CTCACATGGA AAGAAATAAT
2801 GTATTTGTGC CTTCTGTGAG GAATGGGGG AACAAAGTGT CCCAGGTATC
2851 CCCATTTCCA AGGCCCCCCT CCCTCTCCAG GTCCCCCAC AGCAATAAAA
2901 GCTTCCCCCT GATATCCATC CCTTTGTAGT TTGAACAAAT ATATTATAT
2951 GATATGTAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3001 AAA

```

BLAST Results

Entry HS189353 from database EMBL:
human STS WI-11261.
Score = 2191, P = 1.4e-92, identities = 463/485

Medline entries

95310349:
Molecular characterization of NDP52, a novel protein of the nuclear domain 10, which is redistributed upon virus infection and interferon treatment.

97375672:
Cellular localization, expression, and structure of the nuclear dot protein 52.

Peptide information for frame 3

ORF from 57 bp to 2129 bp; peptide length: 691
Category: similarity to known protein
Prosite motifs: RGD (557-560)
LEUCINE_ZIPPER (163-185)
LEUCINE_ZIPPER (475-497)
LEUCINE_ZIPPER (482-504)

```

1 MEESPLSRAP SRGGVNFLNV ARTYIPNTKV ECHYTLPPGT MPSASDWIGI
51 FKVEAACVRD YHTFVWSSVP ESTTDGSPH TSVQFQASYL PKPGAQLYQF
101 RYVNRQGVQC QSPPPFQFRE PRPMDLVTL EEADGGSDIL LVVPKATVLQ
151 NQLDESQQR NDLMQLKLQL EGQVTELR SRVQELERLAT ARQEHTELME
201 QYKGISRSHG EITEERDILS RQQGDHVAR ILEEDDIQTI SEKVLTKVE
251 LDRLRDVTVA LTREQEKL LGKEVQADKE QSEAEQLVAQ QENHHNLNLDL
301 KEAKSQWEEQ SAQAQLKDK VAQMKDTLGQ AQRVALEP LKEQLRGAQE
351 LAASSQKAT LLGEELASAA AARDRTIAEL HRSRLEVAEV NGRLELGLH
401 LKEEKQWSK ERAGLLQSVE AEKDKILKLS AEILRLEKAV QEERTQNQVF
451 KTELAREKDS SLVQLSESKR ELTELRSALR VLQKEKEQLQ EEKQELLEYM
501 RKLEARLEKV ADEKWNEDAT TEDEEAAVGL SCPAALTDSE DESPEDMRP
551 PYGLCERGDG GSSPAGPREA SPLVVISQPA PISPHLSGPA EDSSSDSEAE
601 DEKSVLMAAV QSGGEEANLL LPELGSAFYD MASGFTVGT LSETSTGGPAT
651 PTWKECPICK ERFPAESDKD ALEDHMDGHF FFSTQDPFTF E

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_7p9, frame 3

PIR:A56733 nuclear domain 10 protein NDP52 - human, N = 2, Score = 307,
P = 7.7e-28

TREMBL:AB008852_1 gene: "NDP"; product: "NDP52"; Bos taurus mRNA for NDP52, complete cds., N = 2, Score = 302, P = 4e-27

TREMBL:AC004549_1 gene: "WUGSC:H RG459N13.1"; product: "TXBP151"; Homo sapiens BAC clone RG459N13 from 7p15, complete sequence., N = 2, Score = 275, P = 2.3e-25

PIR:G02043 TXBP151 - human, N = 2, Score = 270, P = 8.5e-25

TREMBL:DM35816 4 gene: "zip"; product: "nonmuscle myosin-II heavy chain"; *Drosophila melanogaster* nonmuscle myosin-II heavy chain (zip) gene, complete cds., N = 1, Score = 254, P = 1.4e-17

>PIR:A56733 nuclear domain 10 protein NDP52 - human
Length = 446

HSPs:

Score = 307 (46.1 bits), Expect = 7.7e-28, Sum P(2) = 7.7e-28
Identities = 104/323 (32%), Positives = 158/323 (48%)

```
Query: 15 VNFLNVARTYIPNTKVECHYTLPPGTMPASDWIGIFKVEAACVRDYHTFVWSSVPESTT 74
      V F +V + YIP V CHYT +P DWIGIF+V R+Y+TF+W ++P
Sbjct: 23 VIFNSVEKFYIPGGDVTCHYTFTQHFIPIRRKDWIGIFRVGWKTTREYYTFMWVTLPIDLN 82

Query: 75 DGSPIHTSVQFQASYLPKPGAQLYQFRYVNRQGVCGQSPFFQFREPRPMDELVTLEED 134
      + S VQF+A YLPK + YQF YV+ G V G S PFQFR D LV +
Sbjct: 83 NKSAKQEQEVQFKAYYLPKDD-EYYQFCYVDEGCVVRGASIPFQFRPENEEIDLVVTTQ-- 139

Query: 135 GGS DILLVVPKATV LQNQ-LDES---QQRNDLMQLKLQLEGQVTE-LRSRVQELERALA 189
      G + + K +NQ L +S Q++N MQ +LQ + + E L+S ++LE +
Sbjct: 140 GEVEEIEQHNLCKENQELKDSCISLQKQNSDMQAE LQKKQEELETLSINKKLELVK 199

Query: 190 TARQE-HTELMEQYKGISRSHEITEERDI-LSRQGGDHVARILELEDDIQTISEKVLTK 247
      + TEL+ Q K ++ E+ I + + Q + E+E +Q +K T+
Sbjct: 200 EQKDYWETELL-QLKEQNQKMSSENEKMGIRVDQLQAQLSTQEKEMEKLVQGDQDK--TE 256

Query: 248 EVE-LDRLRDTVKALTREQEKLGLQKEVQADKEQSEAE LQVAQQENHHLNLDLKEAKSW 306
      ++E L + D + EQ K +L++ +Q+E QQE N DL + S
Sbjct: 257 QLEQLKKENDHFLSLTEQRKDQKKLEQTVEQMKQNETTAMKKQQLMDENFDLSKRLSE 316

Query: 307 QEEQSAQAQRLKDKVAQMMDTLGQAQQRV 335
      E QR K+++ D L + R+
Sbjct: 317 NEIICNALQRQKERLEGENDLLKRENSRL 345
```

Score = 304 (45.6 bits), Expect = 2.1e-27, Sum P(2) = 2.1e-27
Identities = 98/337 (29%), Positives = 163/337 (48%)

```
Query: 15 VNFLNVARTYIPNTKVECHYTLPPGTMPASDWIGIFKVEAACVRDYHTFVWSSVPESTT 74
      V F +V + YIP V CHYT +P DWIGIF+V R+Y+TF+W ++P
Sbjct: 23 VIFNSVEKFYIPGGDVTCHYTFTQHFIPIRRKDWIGIFRVGWKTTREYYTFMWVTLPIDLN 82

Query: 75 DGSPIHTSVQFQASYLPKPGAQLYQFRYVNRQGVCGQSPFFQFREPRPMDELVTLEED 134
      + S VQF+A YLPK + YQF YV+ G V G S PFQFR P +E
Sbjct: 83 NKSAKQEQEVQFKAYYLPKDD-EYYQFCYVDEGCVVRGASIPFQFR---PENE----- 130

Query: 135 GGS DILLVVPKATV LQNQLDESQQERNDLMQLKLQLEGQVTE LRSRVQELERALATARQE 194
      DIL+V Q +++E +Q +L + +L+ L+ + +++ L +QE
Sbjct: 131 --EDILVVT-----QGEVEEIEQHNLCKENQELKDSCISLQKQNSDMQAE LQK-KQE 182

Query: 195 HTELMEQYKGISRSHEITEERDILSRQGGDH-VARILELEDDIQTISEKVLTKVELDR 253
      E ++ I ++ ++ ++Q D+ +L+L++ Q +S + + +D+
Sbjct: 183 ELETLSQ-----INKKLELVKEQKDYWETELLQLKEQNQKMSSENEKMGIRVDQ 232

Query: 254 LRDTVKALTREQEKL--GQLKEVQAD---KEQSEAE LQVAQQENHHLNLDLKEAKSWQE 308
      L+ + +E EKL+ Q K Q + KE L + +Q L+ + Q
Sbjct: 233 LQAQLSTQEKEMEKLVQGDQDKTEQLEQLKKENDHFLSLTEQRKDQKKLEQTVEQMKQN 292

Query: 309 EQSA--QAQRLKDKVAQMMDTLGQAQQRVAELEPLKEQLRGAQEL 351
      E +A + Q L D+ + L + + L+ KE+L G +L
Sbjct: 293 ETTAMKKQQLMDENFDLSKRLSENEIICNALQRQKERLEGENDL 337
```

Score = 124 (18.6 bits), Expect = 2.3e-06, Sum P(2) = 2.3e-06
Identities = 53/227 (23%), Positives = 113/227 (49%)

```
Query: 138 DILLVVPKATV LQNQLDESQQERNDLMQLKLQLEGQVTE LRSRVQELERALATARQEHT 197
      DIL+V Q +++E +Q +L + +L+ L+ + +++ L +QE E
Sbjct: 132 DILVVT-----QGEVEEIEQHNLCKENQELKDSCISLQKQNSDMQAE LQK-KQEELE 185

Query: 198 LMEQYKGISRSHEITEERDILSRQGGDH-VARILELEDDIQTISEKVLTKVELDRDRLD 256
      ++ I ++ ++ ++Q D+ +L+L++ Q +S + + +D+L+
Sbjct: 186 TLQS-----INKKLELVKEQKDYWETELLQLKEQNQKMSSENEKMGIRVDQLQA 235

Query: 257 TVKALTREQEKLGLQKEVQADKEQSEAE LQVAQQENHHLNLDLKEAKSWQE EQSAQAQR 316
      + +E EKL VQ D++++E +L+ ++EN HL L L E + Q++ ++
Sbjct: 236 QLSTQEKEMEKL-----VQGDQDKTE-QLEQLKKENDHFLSLTEQRKDQKKLEQTVEQ 288
```

Query: 317 LK-DKVAQMKDTLGQAQQRVAELEPLKEQLRGAQELA-ASSQKATLLGE 364
 +K ++ MK + Q+ + E L ++L + + A +QK L GE
 Sbjct: 289 MKQNETTAMK----KQELMDENFDLSKRLSENEIICNALQRQKERLEGE 334

Score = 103 (15.5 bits), Expect = 4.4e-04, Sum P(2) = 4.4e-04
 Identities = 63/278 (22%), Positives = 123/278 (44%)

Query: 299 DLKEAKSWQEEQSAQAQRLKDKVAQMK---DTLGQAQQRVAELEPLKEQLRGAQELAAS 354
 +++E + +E + Q LKD ++ D + Q++ ELE L + + EL
 Sbjct: 141 EVEEIEQHNKELCKENQELKQKSDMQAELQKKQEELET--QSINKKLELKV 199

Query: 355 SQQKATLLGEELASAAAARDRTIAELHRSRLVAEVNGRLAELGLHLKEEKCQWSKERAG 414
 Q+ EL + +E + + V ++ +L+ + E+ Q +++
 Sbjct: 200 EQKD--YWETELLQLKEQNQKMSSSENEKMGIRVDQLQAQLSTQEKEM-EKLVQGDQDKTE 256

Query: 415 LLQSVEAEKDKI-LKLSAEIL---RLEKAVQEERTQNVFKTELAREKDSSLVQLSESKR 470
 L+ ++ E D + L L+ + +LE+ V E+ QN+ T +++++ SKR
 Sbjct: 257 QLEQLKKENDHLFLSLTEQRKDQKKLEQTV-EQMKQNET--TAMKKQELMDENFDLSKR 313

Query: 471 ELTELSALRVLQKEKEQLQEEKQELLEYMRKLEARLEKVADEKWNE---DATTEDEEAA 527
 L+E LQ++KE+L+ E +LL ++ +RL +N T DE A
 Sbjct: 314 -LSENEIICNALQRQKERLEGEN-DLL---KRENSRLLSYMGDFNSLPYQVPTSDEGGA 368

Query: 528 ---VGLSCPAALTD-SEDESPEDMRLPPYGLCERGDGSSPAGPREASPL 573
 GL+ + E SP + + +C+ D ++ PL
 Sbjct: 369 RQNPGLAYGNPYSGIQESSSPSLSIKKCPICKADDICDHTLEQQQMQL 418

Score = 64 (9.6 bits), Expect = 7.7e-28, Sum P(2) = 7.7e-28
 Identities = 13/29 (44%), Positives = 17/29 (58%)

Query: 651 PTWKECPICKERFPAESDKDALEDHMDGH 679
 P CPIC + FPA ++K EDH+ H
 Sbjct: 417 PLCFNCPICDKIFPA-TEKQIFEDHVFCH 444

Score = 64 (9.6 bits), Expect = 5.8e+00, Sum P(2) = 1.0e+00
 Identities = 26/90 (28%), Positives = 45/90 (50%)

Query: 470 RELTELSALRVLQKEKEQLQEE---KQELLEYMRKLEARLE-KVADEK--W----- 515
 +E EL+ + LQK+ +Q E KQE LE ++ +LE KV ++K W
 Sbjct: 154 KENQELKQKSDMQAELQKKQEELETLSINKKLELKVKEQKDYWETELLQLK 213

Query: 516 --NEDATTEDEEAAVGLS-CPAALTDSEDE 542
 N+ ++E+E+ + + A L+ E E
 Sbjct: 214 EQNQKMSSSENEKMGIRVDQLQAQLSTQKEKE 243

Score = 47 (7.1 bits), Expect = 4.6e-26, Sum P(2) = 4.6e-26
 Identities = 11/30 (36%), Positives = 17/30 (56%)

Query: 631 MASGFTVGTLSSETSTGGPATPTWKECPICK 660
 +A G + E+S+ P + K+CPICK
 Sbjct: 374 LAYGNPYSGIQESSSPSLSI--KKCPICK 401

Pedant information for DKFZphtes3_7p9, frame 3

Report for DKFZphtes3_7p9.3

[LENGTH] 691
 [MW] 77336.52
 [pI] 4.77
 [HOMOL] PIR:A56733 nuclear domain 10 protein NDP52 - human 2e-29
 [FUNCAT] 09.10 nuclear biogenesis [S. cerevisiae, YDR356w] 2e-11
 [FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YDR356w] 2e-11
 [FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL058w] 2e-11
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YDR356w] 2e-11
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDL058w] 2e-11
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YLR309c] 2e-08
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-07
 [FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-07
 [FUNCAT] 03.25 cytokinesis [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-07
 [FUNCAT] 09.13 biogenesis of chromosome structure [S. cerevisiae, YJL074c] 4e-07
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YNL250w] 4e-06
 [FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YBR289w] 4e-06

[FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YBR289w] 4e-06

[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YBR289w] 4e-06

[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YNL250w] 4e-06

[FUNCAT] 03.13 meiosis [S. cerevisiae, YNL250w] 4e-06

[FUNCAT] 1 genome replication, transcription, recombination and repair [M. jannaschii, MJ1643] 1e-05

[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YJR134c] 4e-05

[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YKR095w] 4e-05

[FUNCAT] 08.19 cellular import [S. cerevisiae, YNL243w] 7e-05

[FUNCAT] 01.03.16 polynucleotide degradation [S. cerevisiae, YNL243w] 7e-05

[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YNL243w] 7e-05

[FUNCAT] 08.99 other intracellular-transport activities [S. cerevisiae, YNL079c] 2e-04

[FUNCAT] 03.01 cell growth [S. cerevisiae, YNL079c] 2e-04

[BLOCKS] BL00682B ZP domain proteins

[EC] 3.6.1.32 Myosin ATPase 1e-13

[PIRKW] nucleus 6e-10

[PIRKW] phosphotransferase 2e-07

[PIRKW] duplication 9e-07

[PIRKW] citrulline 1e-09

[PIRKW] tandem repeat 1e-13

[PIRKW] heart 5e-11

[PIRKW] endocytosis 5e-09

[PIRKW] polymorphism 3e-06

[PIRKW] cornified cell envelope 1e-06

[PIRKW] transmembrane protein 6e-12

[PIRKW] serine/threonine-specific protein kinase 2e-07

[PIRKW] cell wall 1e-06

[PIRKW] zinc finger 5e-09

[PIRKW] metal binding 5e-09

[PIRKW] DNA binding 8e-08

[PIRKW] muscle contraction 1e-11

[PIRKW] IgG constant region-binding 1e-06

[PIRKW] acetylated amino end 4e-09

[PIRKW] actin binding 1e-13

[PIRKW] mitosis 9e-09

[PIRKW] microtubule binding 9e-09

[PIRKW] ATP 1e-13

[PIRKW] thick filament 1e-10

[PIRKW] phosphoprotein 1e-13

[PIRKW] epidermis 1e-06

[PIRKW] leucine zipper 1e-07

[PIRKW] glycoprotein 4e-07

[PIRKW] skeletal muscle 4e-10

[PIRKW] disulfide bond 1e-07

[PIRKW] calcium binding 1e-09

[PIRKW] alternative splicing 1e-10

[PIRKW] coiled coil 1e-13

[PIRKW] P-loop 1e-13

[PIRKW] heptad repeat 6e-10

[PIRKW] methylated amino acid 1e-13

[PIRKW] basement membrane 3e-06

[PIRKW] immunoglobulin receptor 2e-07

[PIRKW] peripheral membrane protein 5e-09

[PIRKW] dimer 1e-07

[PIRKW] cardiac muscle 1e-10

[PIRKW] extracellular matrix 3e-06

[PIRKW] hydrolase 1e-13

[PIRKW] microtubule 6e-10

[PIRKW] muscle 2e-09

[PIRKW] membrane protein 3e-06

[PIRKW] EF hand 1e-09

[PIRKW] cytoskeleton 6e-12

[PIRKW] hair 1e-09

[PIRKW] calmodulin binding 5e-09

[PIRKW] Golgi apparatus 3e-08

[SUPFAM] myosin heavy chain 1e-13

[SUPFAM] conserved hypothetical P115 protein 1e-08

[SUPFAM] hypothetical protein YJL074c 5e-07

[SUPFAM] centromere protein E 9e-09

[SUPFAM] unassigned Ser/Thr or Tyr-specific protein kinases 2e-07

[SUPFAM] calmodulin repeat homology 1e-09

[SUPFAM] myosin motor domain homology 1e-13

[SUPFAM] alpha-actinin actin-binding domain homology 3e-13

[SUPFAM] tropomyosin 3e-07

[SUPFAM] plectin 3e-13

[SUPFAM] trichohyalin 1e-09

[SUPFAM] plectstrin repeat homology 4e-06

[SUPFAM] ribosomal protein S10 homology 3e-13

[SUPFAM] giantin 3e-08
 [SUPFAM] protein kinase homology 2e-07
 [SUPFAM] protein kinase C zinc-binding repeat homology 4e-06
 [SUPFAM] involucrin 1e-06
 [SUPFAM] kinesin motor domain homology 9e-09
 [SUPFAM] human early endosome antigen 1 5e-09
 [SUPFAM] unassigned kinesin-related proteins 8e-08
 [SUPFAM] M5 protein 3e-08
 [SUPFAM] cytoskeletal keratin 3e-08
 [PROSITE] LEUCINE_ZIPPER 3
 [PROSITE] RGD 1
 [PROSITE] MYRISTYL 6
 [PROSITE] CK2_PHOSPHO_SITE 25
 [PROSITE] PKC_PHOSPHO_SITE 6
 [KW] All_Alpha
 [KW] LOW_COMPLEXITY 9.12 %
 [KW] COILED_COIL 39.36 %

SEQ MEESPLSRAPSRGGVNFNLNVRTYIPNTKVECHYTLPPGTMPSASDWIGIFKVEAACVRD
 SEG
 PRD ccc
 COILS
 SEQ YHTFVSSVPESTTDGSPIHSTVQFQASYLPKPGAQLYQFRYVNRQGVCGQSPPFQFRE
 SEG
 PRD eeeeeeeccccccccccccchhhhhhhhhhhhhcccccccccccccccccccccccccccc
 COILS
 SEQ PRPMDELVTLEEADGGSDILLVVPKATVVLQNLDESQQERNLMLQLQLLEGQVTELRSR
 SEG
 PRD cccccceehhhhhchhh
 COILSCC
 SEQ VQELERALATARQEHTELMEQYKGISRSHGEITEERDILSRQGDHVARILELEDDIQT
 SEG
 PRD hhh
 COILS CC
 SEQ SEKVLTKVELDRLRDTVKALTREQEKLGLQKEVQADKEQSEAEQLQVAQQENHHLNLDL
 SEG
 PRD hhh
 COILSCC
 SEQ KEAKSWQEEQSAQAQRLKDKVAQMKDTLGGQAQRVAELEPLKEQLRGAQELAASSQKAT
 SEGxx
 PRD hhh
 COILS CCCCC..CC
 SEQ LLGEELASAAAAARDRTIAELHRSRLVAEVNGLAELGLHLKEKQWSKERAGLLQSVE
 SEG xxxxxxxxxxxxxxxxxxxxxxxx.....
 PRD hhh
 COILS CCCCCCCC.....CCCCCCCCCCCC
 SEQ AEKDKILKLSAEILRLEKAVQEERTQNQVFKTELAREKDSSLVQLSESKRELTELRSLR
 SEG
 PRD hhh
 COILS CC
 SEQ VLQKEKEQLQEEKQELLEMYMRKLEARLEKVADEKWNEDATTEDEEAAGLSCPAALTDSE
 SEG .xxxxxxxxxxxxxxxxxxxxx.....xxxxxxxxxxxxx
 PRD hhh
 COILS CC
 SEQ DESPEDMRLPPYGLCERDGPSSPAGPREASPLVVISQAPAPISPHLSGPAEDSSSDSEAE
 SEGxxxxxxxxxxxxx
 PRD hhhccchh
 COILS
 SEQ DEKSVLMAAVQSGGEEANLLLPELGSAFYDMASGFTVGTLSSETSTGGPATPTWKECPICK
 SEG xx.....
 PRD hhhhhhhhhhhhhcc
 COILS
 SEQ ERFPAESDKDALEDHMDGHFFSTQDPFTFE
 SEG
 PRD cccccccchhhhhhhccccceeecccccccc
 COILS

Prosites for DKFZphtes3_7p9.3

PS00005	190->193	PKC_PHOSPHO_SITE	PDOC00005
PS00005	241->244	PKC_PHOSPHO_SITE	PDOC00005
PS00005	257->260	PKC_PHOSPHO_SITE	PDOC00005
PS00005	468->471	PKC_PHOSPHO_SITE	PDOC00005
PS00005	652->655	PKC_PHOSPHO_SITE	PDOC00005
PS00005	667->670	PKC_PHOSPHO_SITE	PDOC00005
PS00006	28->32	CK2_PHOSPHO_SITE	PDOC00006
PS00006	43->47	CK2_PHOSPHO_SITE	PDOC00006
PS00006	68->72	CK2_PHOSPHO_SITE	PDOC00006
PS00006	72->76	CK2_PHOSPHO_SITE	PDOC00006
PS00006	129->133	CK2_PHOSPHO_SITE	PDOC00006
PS00006	156->160	CK2_PHOSPHO_SITE	PDOC00006
PS00006	208->212	CK2_PHOSPHO_SITE	PDOC00006
PS00006	239->243	CK2_PHOSPHO_SITE	PDOC00006
PS00006	282->286	CK2_PHOSPHO_SITE	PDOC00006
PS00006	305->309	CK2_PHOSPHO_SITE	PDOC00006
PS00006	376->380	CK2_PHOSPHO_SITE	PDOC00006
PS00006	383->387	CK2_PHOSPHO_SITE	PDOC00006
PS00006	468->472	CK2_PHOSPHO_SITE	PDOC00006
PS00006	520->524	CK2_PHOSPHO_SITE	PDOC00006
PS00006	537->541	CK2_PHOSPHO_SITE	PDOC00006
PS00006	539->543	CK2_PHOSPHO_SITE	PDOC00006
PS00006	543->547	CK2_PHOSPHO_SITE	PDOC00006
PS00006	593->597	CK2_PHOSPHO_SITE	PDOC00006
PS00006	595->599	CK2_PHOSPHO_SITE	PDOC00006
PS00006	597->601	CK2_PHOSPHO_SITE	PDOC00006
PS00006	612->616	CK2_PHOSPHO_SITE	PDOC00006
PS00006	639->643	CK2_PHOSPHO_SITE	PDOC00006
PS00006	652->656	CK2_PHOSPHO_SITE	PDOC00006
PS00006	667->671	CK2_PHOSPHO_SITE	PDOC00006
PS00006	683->687	CK2_PHOSPHO_SITE	PDOC00006
PS00008	39->45	MYRISTYL	PDOC00008
PS00008	107->113	MYRISTYL	PDOC00008
PS00008	204->210	MYRISTYL	PDOC00008
PS00008	414->420	MYRISTYL	PDOC00008
PS00008	561->567	MYRISTYL	PDOC00008
PS00008	613->619	MYRISTYL	PDOC00008
PS00016	557->560	RGD	PDOC00016
PS00029	163->185	LEUCINE_ZIPPER	PDOC00029
PS00029	475->497	LEUCINE_ZIPPER	PDOC00029
PS00029	482->504	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphtes3_7p9.3)

DKFZphtes3_8e24

group: signal transduction

DKFZphtes3_8e24.3 encodes a novel 658 amino acid putative GTP-binding protein, related to yeast YGL099w and mouse MMR1 putative GTP-binding proteins.

GTP-binding proteins are involved in various signal transduction pathways, transferring the signal of a cellular receptor to an intracellular signal cascade.

The new protein can find clinical application in modulating/blocking the response to a cellular receptor.

strong similarity to guanine nucleotide binding proteins

complete cDNA, complete cds, potential start at Bp 31, EST hits

Sequenced by MediGenomix

Locus: unknown

Insert length: 3290 bp

Poly A stretch at pos. 3269, polyadenylation signal at pos. 3251

```
1 CGTCCAGCGG TCGTGTGGCC ATGGGCCGGA GGAGAGCCCC GGCCGGTGGG
51 TCCTGGGGAC GGGCCCTTAT GCGCCATCAG ACTCAGCGGA GCCGAAGCCA
101 TCCTCACACT GACTCCTGGT TGCACACAAG TGAACCAAT GATGGCTATG
151 ATTGGGGTCG TCTTAATCTT CAGTCAGTGA CTGAACAGAG CTCCTTGTAT
201 GACTTCCTTG CTACTGCAGA ACTTGCAGGA ACAGAGTTTG TAGCTGAAAA
251 ACTTAATATT AAGTTTGTGC CTGCTGAGGC TAGAAGCTGA CTACTGTCTT
301 TCGAGGAGAG CCAGAGAATT AAGAAGCTCC ATGAAGAAAA CAAACAGTTC
351 TTGTGTATAC CGAGGAGACC AAAGTGGAAC CAAAATACTA CCCGAGAAGA
401 ACTCAACAA GCAGAGAAAG ATAAGTTTCT AGAATGGAGA CGTCAGCTTG
451 TCCGGGTAGA AGAGGAACAG AAGCTGATAT TGACTCCATT TGAACGAAAT
501 TTGGACTTTT GCGCCAGCT CTGGAGAGTC ATTGAGAGAA GTGATATTGT
551 GGTCCAGATA GTAGATGCTC GAAACCCACT CCTGTTTAGA TGTGAGGATT
601 TGGAAATGTTA TGTGAAAGAA ATGGATGCCA ATAAGGAGAA CGTCATTCTG
651 ATCAACAAGG CAGACTTGCT GACTGCTGAG CAGCGGAGTG CCTGGGCCAT
701 GTACTTCGAA AAAGAAGATG TGAAGGTTAT TTTCTGGTCA GCTTTGGCCG
751 GAGCCATTCC CCTGAATGGT GACTCTGAGG AAGAGGCAAA CAGAGATGAT
801 AGACAAAGCA ACACAACTGA GTTGGACAT TCCAGTTTCG ACCAGGCTGA
851 AATTTCCAC AGTGAATCCG AACATCTCCC AGCTAGGGAT TCTCCTTCAC
901 TTAGTGAAAA TCCACAACG GATGAAGATG ACAGTGAGTA TGAGGATGTG
951 CCAGAGGAGG AGGAAGACGA CTGGCAGACG TGCTCAGAAG AAGACGGTCC
1001 CAAGGAAGAG GACTGCAGCC AGGACTGGAA GGAAGCTCT ACTGCAGATT
1051 CTGAGGCTCG GAGCAGGAAA ACCCCACAGA AGAGGCAGAT ACACAATTTT
1101 AGCCATCTGG TATCCAAGCA GGAGTTACTG GAGCTCTTTA AGGAGCTACA
1151 CACTGGGAGA AAGGTGAAAG ATGGCAACT TACGGTCGGA CTGGTGGGCT
1201 ACCCTAATGT TGGTAAGAGT TCAACAATCA ACACCATCAT GGGCAACAAG
1251 AAAGTATCTG TGCTGTCAC ACCTGGTCAC ACAAGCACT TTCAGACTCT
1301 CTATGTGGAG CCTGGCCTCT GCCTGTGTGA CTGCTCTGGC TTGGTGATGC
1351 CATCTTTTGT GTCTACCAAG GCAGAAATGA CTTGCAGCGG AATCCTCCCA
1401 ATTGATCAGA TGAGAGATCA TGTCTCTCCT GTATCACTAG TTTGCCAGAA
1451 TATTTCAAGA CATGTTTTAG AAGCTACCTA TGGCATTAA ATCATAACGC
1501 CTAGAGAGGA TGAAGATCCC CACCGACCTC CAACATCGGA AGAACTGTTG
1551 ACAGCTTATG GATACATGCG AGGATTTCATG ACAGCGCATG GACAGCCAGA
1601 CCAGCCTCGA TCTGCGCGCT ACATCCTGAA GGACTATGTC AGTGGTAAGC
1651 TGCTGTACTG CCATCCTCCT CCTGGAAGAG ATCCTGTAAC TTTTCAGCAT
1701 CAACACCAGC GACTCCTAGA GAACAAAATG AACAGTGATG AAATAAAAAAT
1751 GCAGCTAGGC AGAAATAAAA AAGCAAAGCA GATTGAAAAT ATCGTTGACA
1801 AAACCTTTTT CCATCAAGAG AATGTGAGGG CTTTGACCAA AGGAGTCCAG
1851 GCTGTGATGG GTTACAAGCC CGGGAGTGGT GTAGTGACTG CATCCACTGC
1901 GAGCTCTGAG AACGGGGCGG GGAAGCCCTG GAAAAACAT GGCAACAGAA
1951 ATAAAAAGA AAAAAGTCGT AGACTCTACA AGCACCTGGA TATGTGAGGT
2001 TGGGCTGCAA CAGAAATGTC ATCTGCATTG TGCAGATGGA AAAGAGCAGA
2051 AGCTGCCTGT TGCTGTGGA ACTGTCCCAA GACACTAGCA CTGTAGAACG
2101 GGCCTGTCTC TTGCAGAGCA CGGCTGCACC CAACAGTCTC CATGTCAAGA
2151 CCAAGGGCCT CCTGGAACAA CCAGCTCTGA CAAAAAGGAG TCATCTGGGA
2201 GCCCGAGAAT CTTACTCTG GCCGGGCACA GTGGCTCACG CACCAACATG
2251 GAGAAACCCC GTCTCTACTA AAAATACAAA AAAATTAGCC AGGCGTGGTG
2301 GCGCGCACCT GTAATCCAG CTACTCGGGA GGCTGAGGCA GGGAATCAC
2351 TTGAACCAAG GAGGCAGAGT TTGCAGTGAA TGGAGATTGC GCCGCTGCAC
2401 TCCAGCCTGG GCGACAGAGT GAGACTGCAT CACAAGAAAA AAAATTGCA
2451 AGGGATGGTT CACGAGACAC ATTTGGGACG AAGGTGAAAG AGAAATTCCC
2501 CATCTGAGT GTCTAGTTG GGTCTCTCCG ACTCTAAACA AGGGACTTGG
2551 GTTCAGTTAG TGTACAGCGG GGGCTCACGT CCACTAAGGA ACATGTAGAA
2601 TGTAACCAAC GGGTGACAGG GAAGCTGCGG TATTACTAC CTAGCCCCCA
```

```

2651 TCTTCACTGG TTATTCCACT TATTTAAAT GTCCAGAATA AGCAAATCTC
2701 CATATAGAGG AAGTAGATTA GTGGTTGCTT CGGGATGGGA GGAATGGGAA
2751 GATTGAGGTC TTTCTTTTGC AGTGATAAAA ATGTCCTAAA ATTGACTGTA
2801 GCGATGGTCA CACAACCTCTG AATATGCTTA AGACCATTGA ATTACACACT
2851 TTACGTTGTT GAATTGTATG GTATGTAAAT TATAGTTCAA TAACATAGTT
2901 ACAAAGATA ATCAAAGCA TGAAAGCACT ATTGATGTGG TTTGGATCTG
2951 TGTCTTCACC GAGTCTCATG TTGAAATGTA AGCCCCCTGG TGGGAGGCGA
3001 TGGGATTATG GGGCAGAGTC CTCACAAACG GTTTAGCACC ACCCGCTCAG
3051 TGCTGTTCTC CTGATATTGA GTCCTCATCA CATCTGGTTG CTTCAAAGTG
3101 TGTGGTGCCT CCCCTCTGTC TCCCTCTGTC TCTGGCCATA TAAGATGTGC
3151 CTGCTTCTCC TTCGCCTTCT AACATGATTG TAAGTTTCCT GAGGCCTCCC
3201 TAGAAGCAA AGCTGCTGTG CTCCTGTAC CATCTACTGG ACCGTGAGCC
3251 AATTAAACCT CTTTCTTTA TAAAAAAGG AAAAAAAGG

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 21 bp to 1994 bp; peptide length: 658
 Category: strong similarity to known protein

```

1 MGRRRAPAGG SLGRALMRHQ TQSRSRHRT DSWLHTSELN DGYDWGRNLN
51 QSVTEQSSLD DFLATAELAG TEFVAEKLNI KEVPAEARTG LLSFEESQRI
101 KKLHEENKQF LCIPRRPNWN QNTTPEELKQ AEKDNFLEWR RQLVRLIEEEQ
151 KLILTPFERN LDFWRQLWRV IERSDIVVQI VDARNPLLFR CEDLECYVKE
201 MDANKENVIL INKADLLTAE QRSAWAMYFE KEDVKVIFWS ALAGAIPLNG
251 DSEEEANRDD RQSNTEFGH SSFDQAEISH SESEHLPARD SPSSLSENPTT
301 DEDDSEYEDC PEEEDDWQT CSEEDGPKKE DCSQDWKESS TADSEARSRK
351 TPQKRQIHNF SHLVSKQELL ELFKELHTGR KVKDQQLTVG LVGYPNVGKS
401 STINTIMGNK KVSVSATPGH TKHFQTLYE PGLCLCDCPG LVMPSTFVSTK
451 AEMTCSGILP IDQMRDHVPP VSLVCQNIPR HVLEATYGIN IITPREDEDP
501 HRPPTSEELL TAYGYMRGFM TAHGQPDQPR SARYILKDYV SGKLLYCHPP
551 PGRDPVTFQH QHQRLENKM NSDEIKMQLG RNKKAKQIEN IVDKTFHQE
601 NVRALTKGVQ AVMGYKPGSG VVTASTASSE NGAGKPWKHH GNRNKKESKR
651 RLYKHLDM

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFzptes3_8e24, frame 3

SWISSPROT:YAWG_SCHPO HYPOTHETICAL GTP-BINDING PROTEIN C3F10.16C IN
 CHROMOSOME I., N = 3, Score = 560, P = 1.6e-111

PIR:S64106 hypothetical protein YGL099w - yeast (Saccharomyces
 cerevisiae), N = 2, Score = 544, P = 2.6e-105

TREMBL:CEAF3143_1 gene: "C53H9.2"; Caenorhabditis elegans cosmid
 C53H9., N = 1, Score = 551, P = 2.9e-53

SWISSPROT:MMR1 MOUSE POSSIBLE GTP-BINDING PROTEIN MMR1., N = 2, Score =
 311, P = 7.5e-31

>SWISSPROT:YAWG_SCHPO HYPOTHETICAL GTP-BINDING PROTEIN C3F10.16C IN
 CHROMOSOME I.
 Length = 616

HSPs:

Score = 560 (84.0 bits), Expect = 1.6e-111, Sum P(3) = 1.6e-111
 Identities = 119/253 (47%), Positives = 163/253 (64%)

Query: 12 LGRALMRHQ TQSRSRHRT DSWLHTSELN DGYDWGRNLN QSVTEQSSLD DFLATAELAGT 71
 LGRA+ T+ R+ + H + + R L+SVT ++ LD+FL TAEI

Sbjct: 12 LGRAIQSDFTKNRRNRK--GGLKHIVSDSPKAH--RAALRSVTHETDLDEFNTAELGEV 67

Query: 72 EFVAEKLNIKVP- AEARTGLLSFEESQRIKKLHEENKQFLCIPRRPNWNQNTTPEELKQ 130
EF+AEK N+ + E LLS EE+ R K+ E+NK L IPRRP+W+Q TT EL +

Sbjct: 68 EFIAEQNVTVIQNPQNPFLLSKEEAARSKQKQEKNDRLTIPRRPHWDQTTTAVELDR 127

Query: 131 AEKDNFLEWRRQLVRLEEEQKLILTPFERNLDFWRQLWRVIERSDIVVQIVDARNPLLFR 190
E+++FL WRR L +L++ + I+TPFERNL+ WRQLWRVIERSD+VVQIVDARNPL FR

Sbjct: 128 MERESFLNWRRLAQLQDVEGFIVTPFERNLEIWRQLWRVIERSDVVVQIVDARNPLFFR 187

Query: 191 CEDLECYVKEMDANKENVILINKADLLTAEQSAWAMYFEKEDVKVIFWSALAGAIPLNG 250
LE YVKE+ +K+N +L+NKAD+LT EQR+ W+ YF + ++ +F+SA A N

Sbjct: 188 SAHLEQYVKEVGPSKKNFLLVNKADMLTEEQRNYWSSYFNENNIPFLFFSARMAA-EANE 246

Query: 251 DSEEEANRDDRQSN 264
E+ + SN

Sbjct: 247 RGEDLETYESTSSN 260

Score = 532 (79.8 bits), Expect = 1.6e-111, Sum P(3) = 1.6e-111
Identities = 131/323 (40%), Positives = 192/323 (59%)

Query: 340 STADSEARSRKTPQKRQIHNFSLVSKQELLEFLKELHTGRKVKDGG--LTVGLVGYPNV 397
ST+ +E + +H+ S + + + L +F++ + + DG+ +T GLVGYPNV

Sbjct: 256 STSSNEIPESLQADENDVHS-SRIATLKVLEGIFEKFFAS--TLPDGKTKMTFGLVGYPNV 312

Query: 398 GKSSTINTIMGNKKVSVSATPGHTKHFQTLVVEPGLCLDCPGLVMPSFVSTKAEMTCSG 457
GKSSTIN ++G+KKVSVS+TPG TKHFQT+ + + L DCPGLV PSF +T+A++ G

Sbjct: 313 GKSSTINALVGSKKVSVSSTPGTKKHFQTINLSEKVSLLDCPGLVFPFATTQADLVLDG 372

Query: 458 ILPIDQMRDHVPVSLVCQNI PRHVLEATYGINI-ITPREDEDPHRPPTSEELLTAYGYM 516
+LPIDQ+R++ P +L+ + IP+ VLE Y I I I P E E P+++E+L +

Sbjct: 373 VLPIDQLREYTGPSALMAERIPKEVLETLYTIRIRIKPIE-EGGTGVPQAQEVLPFFARS 431

Query: 517 RGEMTAH-GQPDQPSARYILKDYVSGKLLYCHPPPG--RDPVTFQHQHQRLLNKMNSD 573
RGEM AH G PD R+AR +LKDYV+GKLLY HPPP F +H + + + SD

Sbjct: 432 RGEMRAHHGTPDDSRARILLKDYVNGKLLYVHPPPNYPNSGSEFNKEHHQKIVSA-TSD 490

Query: 574 EIKMQLGR---NKKAKQIEN-IVDKTFFHQEN--VRALTGKVQAVM-G--YKPGSGVVT 624
I +L R + E+ +VD +F QEN VR + KG M G YK + +

Sbjct: 491 SITEKQRTAISDNLTLSAESQLVDDEYF-QENPHVRPMVKGTA VAMQGPVYKGRNTMQPF 549

Query: 625 STASSENGAGK-PWKKHGNRNKKEKSRL 652
+++ + K P G + K+R+L

Sbjct: 550 QRRNLDDASPKYPMNAQKGKPLSRRKARQL 578

Score = 47 (7.1 bits), Expect = 1.3e-60, Sum P(3) = 1.3e-60
Identities = 21/84 (25%), Positives = 35/84 (41%)

Query: 552 GRDPVTFQHQHQRLLNKMNSDEIKMQLGRNKKAKQIENIVDKTFFHQENVRALTGKVQA 611
G D T++ + + +DE + R K +E I +K F TK

Sbjct: 248 GEDLETYESTSSNEIPESLQADENDVHSSRIATLKVLEGIFEK--FASTLPDGKTKMTFG 305

Query: 612 VMGYKPGSGVVTASTASSENGAGK 635
++GY P G +ST ++ G+ K

Sbjct: 306 LVGY-PNVG--KSSTINALVGSKK 326

Score = 43 (6.5 bits), Expect = 1.6e-111, Sum P(3) = 1.6e-111
Identities = 7/13 (53%), Positives = 9/13 (69%)

Query: 638 KKHGNRNKKEKS 650
KKH +NK+ K R

Sbjct: 596 KKHNNKRSKQR 608

Pedant information for DKFZphtes3_8e24, frame 3

Report for DKFZphtes3_8e24.3

[LENGTH] 658
[MW] 75226.58
[PI] 5.86
[HOMOL] SWISSPROT:YAWG_SCHPO HYPOTHETICAL GTP-BINDING PROTEIN C3F10.16C IN CHROMOSOME
I. 5e-56
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YGL099w] 3e-55
[FUNCAT] r general function prediction [M. jannaschii, MJ1464] 1e-16
[FUNCAT] 08.16 extracellular transport [S. cerevisiae, YER006w] 3e-09
[PIRKW] P-loop 1e-27
[PIRKW] GTP binding 1e-27
[SUPFAM] conserved hypothetical protein MG442 7e-08

```

[PROSITE]      ATP_GTP_A      1
[PROSITE]      MYRISTYL      3
[PROSITE]      AMIDATION      2
[PROSITE]      CAMP_PHOSPHO_SITE      1
[PROSITE]      CK2_PHOSPHO_SITE      19
[PROSITE]      TYR_PHOSPHO_SITE      2
[PROSITE]      PKC_PHOSPHO_SITE      10
[PROSITE]      ASN_GLYCOSYLATION      2
[KW]           Alpha_Beta
[KW]           LOW_COMPLEXITY      4.56 %

SEQ      MGRRRAPAGGSLGRALMRHQTRSRSHRHTDSWLHTSELNDGYDWGRNLQSVTEQSSLD
SEG      .....XXXXXXXXXXXXX.....
PRD      cccccccccchhhhhhhhhhhccccccccccccccccccccchhhhhhhhhccccch

SEQ      DFLATAELAGTEFVAEKLNIKFPVPAEARTGLLSFEESQRIKKLHEENKQFLCIPRRPNWN
SEG      .....
PRD      hhhhhhhhhhhheeeccccceeeeeeccccchhhhhhhhhhhhhhhhhhhhhcccccccc

SEQ      QNTTPEELKQAEKDNFLEWRRLVRLEEEQKLILTPFERNLDFWRQLWRVIERSDIVVQI
SEG      .....
PRD      cccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhccccceeee

SEQ      VDARNPLLFRCEDLECYVKEMDANKENVILINKADLLTAEQRSAWAMYFEKEDVKVIFWS
SEG      .....
PRD      eccccccccchhhhhhhhhhhccccceeeeeeccccchhhhhhhhhhhhhhhhhhhccccceeeec

SEQ      ALAGAIPLNGDSEEEANRRDRQSNTEFGHSSFDQAEISHSESEHLPARDSPSLSENPTT
SEG      .....
PRD      cccccccccchhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccc

SEQ      DEDDSEYEDCPREEEDDQWTCSEEDGPKEDCSQDWKESSTADSEARSRKTPOKRQIHNF
SEG      .....
PRD      cccccccccccccccccccccccccccccccccccccccccchhhhhhhhhcccccccccccc

SEQ      SHLVSKQELLELFKELHTGRKVKDQGLTVGLVGYPNVGKSSINTIMGNKKVSVSATPGH
SEG      .....
PRD      cccccchhhhhhhhhhhhhhhhhhhccccceeeeeeccccccccccccceeeccccceeecccccc

SEQ      TKHFQTLYVEPGLCLDCPGLVMPFSFVSTKAEMTCSGILPIDQMRDHVPPVSLVCQNIPR
SEG      .....
PRD      cceeeeeeccccceccccccccccccchhhhhhhhhccccccccccccccccceeeccccch

SEQ      HVLEATYGINIITPREDEDPHRPPTSEELLTAYGYMRGFMHTAGQPDPQRSARYILKDYV
SEG      .....
PRD      hhhhhhhhhccccccccccccccccchhhhhhhhhhhhhhhhhccccccccchhhhhhhhhhhcc

SEQ      SGKLLYCHPPGRDPVTFQHQHQRLENKMNSDEIKMQLGRNKKAKQIENIVDKTFFHQE
SEG      .....
PRD      cceeeccccccccccccchhhhhhhhhhhccccchhhhhhhhhccccchhhhhhhhhhhccccch

SEQ      NVRALTKGVQAVMGYKPGSGVVTASTASSENGAGKPWKKHGNNRNNKKEKSRRLYKHLDM
SEG      .....
PRD      hhhhhhhceeeeeeccccceeeccccccccccccccccccccccccchhhhhhhhhhhccc

```

Prosites for DKFZphtes3_8e24.3

PS00001	264->268	ASN_GLYCOSYLATION	PDOC00001
PS00001	359->363	ASN_GLYCOSYLATION	PDOC00001
PS00004	410->414	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	21->24	PKC_PHOSPHO_SITE	PDOC00005
PS00005	26->29	PKC_PHOSPHO_SITE	PDOC00005
PS00005	97->100	PKC_PHOSPHO_SITE	PDOC00005
PS00005	348->351	PKC_PHOSPHO_SITE	PDOC00005
PS00005	378->381	PKC_PHOSPHO_SITE	PDOC00005
PS00005	448->451	PKC_PHOSPHO_SITE	PDOC00005
PS00005	493->496	PKC_PHOSPHO_SITE	PDOC00005
PS00005	531->534	PKC_PHOSPHO_SITE	PDOC00005
PS00005	541->544	PKC_PHOSPHO_SITE	PDOC00005
PS00005	649->652	PKC_PHOSPHO_SITE	PDOC00005
PS00006	52->56	CK2_PHOSPHO_SITE	PDOC00006
PS00006	57->61	CK2_PHOSPHO_SITE	PDOC00006
PS00006	93->97	CK2_PHOSPHO_SITE	PDOC00006
PS00006	123->127	CK2_PHOSPHO_SITE	PDOC00006
PS00006	155->159	CK2_PHOSPHO_SITE	PDOC00006
PS00006	252->256	CK2_PHOSPHO_SITE	PDOC00006
PS00006	271->275	CK2_PHOSPHO_SITE	PDOC00006
PS00006	279->283	CK2_PHOSPHO_SITE	PDOC00006

PS00006	281->285	CK2_PHOSPHO_SITE	PDOC00006
PS00006	293->297	CK2_PHOSPHO_SITE	PDOC00006
PS00006	299->303	CK2_PHOSPHO_SITE	PDOC00006
PS00006	305->309	CK2_PHOSPHO_SITE	PDOC00006
PS00006	320->324	CK2_PHOSPHO_SITE	PDOC00006
PS00006	322->326	CK2_PHOSPHO_SITE	PDOC00006
PS00006	340->344	CK2_PHOSPHO_SITE	PDOC00006
PS00006	365->369	CK2_PHOSPHO_SITE	PDOC00006
PS00006	449->453	CK2_PHOSPHO_SITE	PDOC00006
PS00006	493->497	CK2_PHOSPHO_SITE	PDOC00006
PS00006	505->509	CK2_PHOSPHO_SITE	PDOC00006
PS00007	480->488	TYR_PHOSPHO_SITE	PDOC00007
PS00007	190->198	TYR_PHOSPHO_SITE	PDOC00007
PS00008	9->15	MYRISTYL	PDOC00008
PS00008	432->438	MYRISTYL	PDOC00008
PS00008	620->626	MYRISTYL	PDOC00008
PS00009	1->5	AMIDATION	PDOC00009
PS00009	378->382	AMIDATION	PDOC00009
PS00017	393->401	ATP_GTP_A	PDOC00017

(No Pfam data available for DKFZphtes3_8e24.3)

DKFZphtes3_8g11

group: testes derived

DKFZphtes3_8g11 encodes a novel proline-rich 939 amino acid protein without similarity to known proteins.

The novel protein contains an ATP/GTP-binding site motif A (P-loop).
No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown, prolin ritch protein

1 EST hit (from testis library)

Sequenced by MediGenomix

Locus: unknown

Insert length: 3100 bp

Poly A stretch at pos. 3056, polyadenylation signal at pos. 3041

```
1 AGAGTCTTCC CTCAGCATAT TTTACGATAG AGAAGATCTT GTTCCAATGG
51 AAGAAAGTGA GGACTCACAG AGTGATTCCC AGACAAGGAT TTCTGAGTCC
101 CAACACTCCC TCAAGCCAAA TTATCTTTCC CAGGCCAAGA CTGACTTCTC
151 AGAACAGTTC CAGTTGCTAG AAGATCTGCA GCTAAAAATA GCAGCAAAAC
201 TCTTAAGGAG TCAAATACCC CCCGATGTGC CTCCACCTCT AGCTTCAGGT
251 CTAGTCCTAA AATACCCTAT CTGCCTACAG TGTGGCCGAT GTTCAGGACT
301 TAATTGCCAT CATAAAATTAC AGACCACCTT GGGGCCCTTAT CTTCCTTATCT
351 ATCCACAGCT CCACCTTGTA CGCACTCCTG AAGGCCATGG TGAGGTTCCGG
401 TTGCATCTTG GCTTTAGGCT GAGAATTGGG AAAAGATCCC AAATCTCAAA
451 GTATCGTGAA AGAGATAGAC CCGTCATACG GAGAAGCCCT ATATCACCAT
501 CACAAAAGGAA AGCTAAAATC TATACTCAAG CTTCACAGAG TCCTACTTCC
551 ACAATAGATT TGCAGTCTGG GCCTTCCAG TCCCCTGCTC CTGTACAAGT
601 CTACATCAGG CGAGGACAAC GCAGCAGGCC TGACTTAGTA GAAAAGACAA
651 AAACCTAGAGC ACCTGGGCAC TATGAATTCA CTCAAGTTCA CAACCTACCA
701 GAGAGTGACT CTGAAAAGCAC TCAGAATGAA AAACGGGCTA AAGTGAGAAC
751 CAAAAAGACC TCTGATTCAA AATATCCAAT GAAGAGAATC ACCAAGCGAC
801 TTAGAAAACA CAGAAAGTTC TACACAAACA GTAGAACCAC AATAGAGAGT
851 CCTTCTAGGG AATTAGCAGC CCATTTAAGA AGGAAGAGGA TTGGAGCAAC
901 TCAGACAAGT ACTGCCCTCT TAAAAAGACA ACCTAAGAAA CCTTCCCAAC
951 CCAAGTTTCAT GCAACTGCCT TTTAGAGGCC TAAAGCGGGC ATTCCAAACA
1001 GCACACAGAG TTATAGCTTC TGTGGGGCGG AAGCCTGTGG ACGGGACAAG
1051 GCCAGACAAT TTGTGGGCAA GCAAAAACCTA TTATCCAAAA CAAAATGCCA
1101 GGGACTATTG CTTACCAAGC AGTATCAAAA GAGACAAGAG GTCAGCTGAC
1151 AAGCTAACGC CAGCAGGCTC AACCATTAA GAGGAGGACA TATTGTGGGG
1201 AGGAACGGTC CAGTGCAGAT CAGCTCAACA GCCAAGAAGA GCTTACTCTT
1251 TCCAACCCAG ACCTCTTCGA CTGCCCAAGC CCACAGATT CCAAAGTGGT
1301 ATTGCTTTCC AAACCTGCCTC AGTGGGGCAG CCTCTGAGAA CTGTTCAAAA
1351 GGACAGTAGT AGCAGATCAA AGAAAACTT CTATAGAAAT GAAACCTCCA
1401 GCCAGGAGTC TAAGAACTTG TCCACACAGG GAACAGAGT TCAGGCCCGA
1451 GGAAGAATCC TACCTGGTTC CCCTGTGAAG AGAACCTGGC ACCGACATCT
1501 TAAAGACAAA CTCACACACA AGGAGCATAA CCACCCAGC TTCTATAGGG
1551 AGAGAACCCC ACGCGGTCCT TCTGAGAGAA CCCGTCATAA CCCCTCTTGG
1601 AGAAACCATC GCAGTCCCTC TGAGAGAAGC CAACGCAGTT CCTTGGAGAG
1651 AAGACATCAC AGTCCCTCTC AGAGGAGCCA CTGCAGTCCC TCTAGGAAAA
1701 ACCATTCCAG TCCTTCTGAG AGAAGCTGGC GCAGTCCGTC TCAGAGAAAT
1751 CACTGCAGTC CCCCAGAGAG GAGCTGTAC AGTCTCTCTG AAAGGGGCCT
1801 TCACAGTCCC TCTCAGAGGA GCCATCGCGG TCCCTCTCAG AGAAGACATC
1851 ACAGTCCCTC AGAGAGAAGC CATCGCAGTC CCTCAGAGAG AAGCCATCGC
1901 AGTCCCTCTG AGAGAAGACA TCGCAGTCCC TCCCAGAGGA GCCATCGCGG
1951 TCCTCAGAG AGAAGCCATT GCAGTCCCTC TGAGAGAAGA CATCGCAGTC
2001 CCTCTCAGAG GAGCCATCGT GGTCCCTCTG AGAGAAGACA TCACAGTCCC
2051 TCTAAGAGAA GCCATCGCAG TCCCGCTCGG AGGAGCCATC GCAGTCCCTC
2101 AGAGAGAAGC CATCACAGTC CCTCTGAGAG AAGCCATCAC AGTCCCTCTG
2151 AGAGAAGACA TCACAGTCCC TCTGAGAGAA GCCATTGCAG TCCCTCTGAG
2201 AGAAGCCATT GCAGTCCCTC TGAGAGAAGA CATCGCAGTC CCTCTGAGAG
2251 AAGACATCAC AGTCCCTCAG AGAAAAGCCA TCACAGTCCC TCTGAGAGAA
2301 GCCATCACAG TCCCTCTGAG AGAAGACGTC ACAGTCCCTT GGAGAGGAGC
2351 CGTCACAGTC TCTTGGAGAG GAGCCATCGC AGTCCCTCTG AGAGGAGATC
2401 TCACAGGTCC TTTGAGAGGA GCCATCGTAG GATTCTCTGAG AGAAGTCACA
2451 GTCCCTCAGA GAAGAGCCAC CTCAGTCCCT TGGAAGAAG CCGTTGCAGT
2501 CCCTCTGAGA GGAGAGGACA CAGTTCCTCT GGGAAAACCT GTCACAGTCC
2551 CTCTGAGAGA AGCCATCGCA GTCCCTCCGG GATGAGGCAA GGGAGGACCT
2601 CTGAGAGGAG CCATCGCAGT TCCTGTGAGA GAACCCGTCA CAGTCCCTCT
```

```

2651 GAGATGAGGC CAGGGAGGCC CTCTGGGAGG AACCATGCA GTCCCTCTGA
2701 GAGGAGCCGA CGCAGTCCCC TTAAGGAGGG ACTCAAGTAC AGTTTCCTG
2751 GAGAGAGGCC CAGCCATAGT TTGTCTAGAG ATTTCAAGAA TCAAACAACT
2801 CTCCTCGGGA CCACACATAA AAATCCCAAA GCAGGGCAAG TGTGGAGGCC
2851 TGAAGCTACT CGATGAGGCG AGGTCCGCCC CTATTATTCA TTGTCTTAAG
2901 TCTTCATCGT GCTGCCCTTT CCAGGCTTCT TTCCTGCTCA GCCACTGCCT
2951 CCAATTCCTG CGCCCCCAGC GTGGAAGGC TTCCATTCT CTCTACCGGG
3001 GGGGAGCGG GTGAGAAATG GTCTGTAATT TCTCTAAGAT GAATAAAGGG
3051 GCAGTTAATT AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAGG

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 47 bp to 2863 bp; peptide length: 939
 Category: similarity to unknown protein
 Classification: unclassified
 Prosite motifs: ATP_GTP_A (824-832)

```

1 MEESEDSQSD SQTRISESQH SLKPNYLSQA KTDSEQFQL LEDLQLKIAA
51 KLLRSQIPPD VPPPLASGLV LKYPICLQCG RCGSLNCHHK LQTTSGPYLL
101 IYPQLHLVRT PEGHGEVRLH LGFRLRIGKR SQISKYRERD RPIVRRSPIS
151 PSQRKAKIYT QASKSPTSTI DLQSGPSQSP APVQVYIRRG QSRPDLVEK
201 TKTRAPGHYE FTQVHNLPEP DSESTQNEKR AKVRTKKTSD SKYPMKRITK
251 RLRKHKRFYT NSRTTIESPS RELAAHLRRK RIGATQTSTA SLKRQPKKPS
301 QPKFMQLLFQ SLKRAFQTAH RVIASVGRKP VDGTRPONLW ASKNYYPKQN
351 ARDYCLPSSI KRDKRSADKL TPAGSTIKQE DILWGGTVQC RSAQQPRRAY
401 SFQPRPLRLP KPTDSQSGIA FQTASVGQPL RTVQKDSSSR SKKNFYRNET
451 SSQESKNLST PGTRVQAGR ILPGSPVKRT WHRHLKDKLT HKEHNHPSFY
501 RERTPRGPSE RTRHNPSWRN HRSPPERSQR SSLERRHHSP QSRSHCSPSR
551 KNHSSPSERS WRSPPSQRNHC SPPERSCHSL SERGLHSPSQ RSHRGPSQRR
601 HHSPPERSHR SPSPERSHRP SERRHRSRSP RSHRGPSERS HCSPPERRHR
651 SPSQRSHRGP SERRHHSPPSK RSHRSPARRS HRSPPERSHH SPSPERSHHSP
701 SERRHHSPPSE RSHCSPERS HCSPPERRHR SPSPERRHSP SEKSHHSPSE
751 RSHHSPSERR RSHPLERSRH SLLERSHRSP SERRSHRSFE RSHRRISERS
801 HSPSEKSHLS PLERSRCSPPS ERGRHSSSGK TCHSPERSRH RSPSGMRQGR
851 TSERSHRSSC ERTRHSPSEM RPRGRPSGRNH CSPSERSRRS PLKEGLKYSF
901 PGERPSHSLR RDFKNQTTLL GTTHKNPKAG QVWRPEATR

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFzphes3_8g11, frame 2

TREMBL:AF061185_1 gene: "car90"; product: "cyst germination specific acidic repeat protein precursor"; *Phytophthora infestans* cyst germination specific acidic repeat protein precursor (car90) gene, complete cds., N = 1, Score = 457, P = 2.3e-39

TREMBL:AC004561_38 gene: "F16P2.41"; product: "putative proline-rich protein"; *Arabidopsis thaliana* chromosome II BAC F16P2 genomic sequence, complete sequence., N = 1, Score = 340, P = 4.2e-27

TREMBL:AF062655_1 product: "plenty-of-prolines-101"; *Mus musculus* plenty-of-prolines-101 mRNA, complete cds., N = 1, Score = 313, P = 3.6e-24

PIR:PN0099 son3 protein - human (fragment), N = 1, Score = 292, P = 1.2e-22

>TREMBL:AF061185_1 gene: "car90"; product: "cyst germination specific acidic repeat protein precursor"; *Phytophthora infestans* cyst germination specific acidic repeat protein precursor (car90) gene, complete cds.

Length = 1,489

HSPs:

Score = 457 (68.6 bits), Expect = 2.3e-39, P = 2.3e-39
 Identities = 91/444 (20%), Positives = 239/444 (53%)

Query: 475 SPVKRTWHRHLKDKLTHKEHNHPSFY-RERTPRGPSETRHNPSWRNHRSPSERSQSSSL 533
 +P + T + +++ T+ ++ E TP P+E T + P+ +P+E + +S
 Sbjct: 584 APTEETMYAPIET-TYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAST 642

Query: 534 ERRHHSQSRSHCSPSRKNHSSPSERSWRSPSQRNHCSPPERSCHSLSERGLHSPSQSRSH 593
 E ++P++ + +P+ + P+E + +P++ +P E + ++ +E ++P++ +
 Sbjct: 643 EETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETT 702

Query: 594 RGPQRHHSERSHRSPSERSHRSPSERRHRSPSQSRHSGPSERSHCSPSERRHRSPS 653
 P++ + P+E + +P+E + +P+E +P + + GP+E + +P+E +P+
 Sbjct: 703 YAPAEETPYEPTTEETTYAPTEETTYAPTEETMYAPIETTYGPTTEETTYAPTEETTYAPT 762

Query: 654 QSRHSGPSERRHHSPSKRSHRSPARRSHRSPSERSHHSERSHHSERSHHSERSH 713
 + + P+E + P+ + +P + +P+E + ++P+E + ++P+E + P+E +
 Sbjct: 763 EETTYAPTEETTYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETT 822

Query: 714 CSPSERSHCSPSERRHRSPSERRHHSPEKSHHSPSERSHHSERSHHSERSHHSPLERSHLL 773
 +P+E + P+E +P+E ++P+E++ ++P+E++ ++P+E ++P E + +
 Sbjct: 823 YAPTEETPYEPTTEETTYTPTTEETTYAPTEETTYAPTEKTYAPTEETTYAPTEETPYEPT 882

Query: 774 ERSRSPSERRSHRSFERS-HRRISERSHSPSEKSHLSPLERSRCSPSERRGHSSSGKTC 832
 E + +P++ ++ E + + E +++P+E++ +P E + P+E ++ + +T
 Sbjct: 883 EETTYAPTKETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETT 942

Query: 833 HSPSERSHRSPSGMRQGRTERSRRSSCERTRHSPSEMRRPGRPSGRNHCSPSERSRRSPL 892
 ++P+E + +P+ +E + + E T + P+E P+ +P+E + +P+
 Sbjct: 943 YAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETMYAPI 1002

Query: 893 KEGLKYSFPGERPSHLSRDFKNQTT 918
 +E Y+ P E +++ + + + T
 Sbjct: 1003 EE-TTYA-PTEETTYAPAEETPYEPT 1026

Score = 445 (66.8 bits), Expect = 4.5e-38, P = 4.5e-38
 Identities = 83/394 (21%), Positives = 212/394 (53%)

Query: 502 ERTPRGPSETRHNPSWRNHRSPSERSQSSSLERRHHSQSRSHCSPSRKNHSSPSERSW 561
 E TP P+E T + P+ +P+E + + E ++P++ + +P+ + P+E +
 Sbjct: 763 EETTYAPTEETTYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTTEETT 822

Query: 562 RSPSQRNHCSPPERSCHSLSERGLHSPSQSRHSGPSQRRHHSPSERSHRSPSERSHRSPS 621
 +P++ P E + ++ +E ++P++ + P+++ ++P+E + +P+E + P+
 Sbjct: 823 YAPTEETPYEPTTEETTYTPTTEETTYAPTEETTYAPTEKTYAPTEETTYAPTEETPYEPT 882

Query: 622 ERRHRSQSRHSGPSERSHCSPSERRHRSPSQSRHSGPSERRHHSPSKRSHRSPARRSH 681
 E +P++ + P+E + + +E +P++ + P+E + P++ + +P +
 Sbjct: 883 EETTYAPTKETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETT 942

Query: 682 RSPSERSHHSERSHHSPSERRHHSPSERSHCSPSERSHCSPSERRHRSPSERRHHSPS 741
 +P+E + ++P+E + ++P+E ++P+E + P+E + +P+E +P+E ++P
 Sbjct: 943 YAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETMYAPI 1002

Query: 742 EKSHHSPSERSHHSPSERRHHSPLERSRHSLLERSHRSPSERRSHRSFERS-HRRISERS 800
 E++ ++P+E + ++P+E + P E + ++ E + +P+E ++ S E + + E +
 Sbjct: 1003 EETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYASTEETTYAPTEETT 1062

Query: 801 HSPSEKSHLSPLERSRCSPSERRGHSSSGKTCSPSERSHRSPSGMRQGRTERSRRSSC 860
 ++P+E++ P E + +P+E ++ + +T ++P+E + +P+ +E +
 Sbjct: 1063 YAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPT 1122

Query: 861 ERTTRHSPSEMRRPGRPSGRNHCSPSERSRRSPLKE 894
 E T ++P+E P+ +P E + P +E
 Sbjct: 1123 EETTYAPTEETTYAPTEETMYAPIETTYGPTTE 1156

Score = 439 (65.9 bits), Expect = 2.0e-37, P = 2.0e-37
 Identities = 86/421 (20%), Positives = 223/421 (52%)

Query: 475 SPVKRTWHRHLKDKLTHKEHNHPSFY-RERTPRGPSETRHNPSWRNHRSPSERSQSSSL 533
 +P + T + +K T+ ++ E TP P+E T + P+ +P+E + +S
 Sbjct: 848 APTEETTYAPT-EKTTYAPTEETTYAPTEETPYEPTTEETTYAPTKETTYAPTEETTYAST 906

Query: 534 ERRHHSQSRSHCSPSRKNHSSPSERSWRSPSQRNHCSPPERSCHSLSERGLHSPSQSRSH 593
 E ++P++ + +P+ + P+E + +P++ +P E + ++ +E ++P++ +
 Sbjct: 907 EETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETT 966

Query: 594 RGPSQRRHHSRSPERSHRSPERSHRSPERRHRSRSPQRSHRGPSERSHCSRSPERRHRSRSPS 653
 P++ + P+E + +P+E + +P+E +P + + P+E + +P+E P+
 Sbjct: 967 YAPAEETPYEPTTEETTYAPTEETTYAPTEETMYAPIEETTYAPTEETTYAPAEETPYEPT 1026

Query: 654 QRSHRGPSERRHHSRSPKSRSHRSPARRSHRSPERSHHSRSPERSHHSRSPERRHHSRSPERSH 713
 + + P+E ++P++ + + + +P+E + ++P+E + + P+E ++P+E +
 Sbjct: 1027 EETTYAPTEETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETT 1086

Query: 714 CSPSERSHCSRSPERRHRSRSPERRHHSRSPSEKSHHSRSPERSHHSRSPERRHHSPLERSRHSLL 773
 +P+E + +P+E +P+E ++P+E++ + P+E + ++P+E ++P E + ++ +
 Sbjct: 1087 YAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETMYAPI 1146

Query: 774 ERSRSPSERRSHRSFERS-HRRISERSHSPSEKSHLSPLERSRCSRSPERRGHSSSGKTC 832
 E + P+E ++ E + + E ++P+E++ P + +P+E ++ + +T
 Sbjct: 1147 EETTYGPTTEETTYAPTEATTYAPTEETPYAPTEETTYEPTGETTYAPTEETTYAPTEETT 1206

Query: 833 HSPSERSHRSPSGMRQGRTERSRRSSCERTRHSPSEMRPGRPSGRNHCSRSPERSRRSPL 892
 ++P+E + +P+ +E + + E T + P+E P+ +P+E + +P
 Sbjct: 1207 YAPTEETTYAPTEETPYEPTTEETTYAPTEETTYEPTTEETTYAPTEETTYAPTEETTYAPT 1266

Query: 893 KE 894
 +E
 Sbjct: 1267 EE 1268

Score = 439 (65.9 bits), Expect = 2.0e-37, P = 2.0e-37
 Identities = 91/434 (20%), Positives = 232/434 (53%)

Query: 475 SPVKRTWHRHLKDKLTHKEHNHPSFY-RERTPRGPSETRHNPSWRNHRSRSPERSQRSSL 533
 +P + T + +K T+ ++ E TP P+E T + P+ +P+E + +S
 Sbjct: 440 APTEETTYAPT-EKTTYAPTEETTYAPTEETPYEPTTEETTYAPTKETTYAPTEETTYAST 498

Query: 534 ERRHHSRSPQRSHCSRSPKRNHSSPERSWRSPSRQNRHCSPPERSCHLSERGLHSPSRSH 593
 E ++P++ + +P+ + P+E + +P++ +P E + ++ +E ++P++ +
 Sbjct: 499 EETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETT 558

Query: 594 RGPSQRRHHSRSPERSHRSPERSHRSPERRHRSRSPQRSHRGPSERSHCSRSPERRHRSRSPS 653
 P++ + P+E + +P+E + +P+E +P + + P+E + +P+E P+
 Sbjct: 559 YAPAEETPYEPTTEETTYAPTEETTYAPTEETMYAPIEETTYAPTEETTYAPAEETPYEPT 618

Query: 654 QRSHRGPSERRHHSRSPKSRSHRSPARRSHRSPERSHHSRSPERSHHSRSPERRHHSRSPERSH 713
 + + P+E ++P++ + + + +P+E + ++P+E + + P+E ++P+E +
 Sbjct: 619 EETTYAPTEETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETT 678

Query: 714 CSPSERSHCSRSPERRHRSRSPERRHHSRSPSEKSHHSRSPERSHHSRSPERRHHSPLERSRHSLL 773
 +P+E + +P+E +P+E ++P+E++ + P+E + ++P+E ++P E + ++ +
 Sbjct: 679 YAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETMYAPI 738

Query: 774 ERSRSPSERRSHRSFERS-HRRISERSHSPSEKSHLSPLERSRCSRSPERRGHSSSGKTC 832
 E + P+E ++ E + + E ++P+E++ P + +P+E ++ + +T
 Sbjct: 739 EETTYGPTTEETTYAPTEATTYAPTEETPYAPTEETTYEPTGETTYAPTEETTYAPTEETT 798

Query: 833 HSPSERSHRSPSGMRQGRTERSRRSSCERTRHSPSEMRPGRPSGRNHCSRSPERSRRSPL 892
 ++P+E + +P T E + + E T ++P+E P P+ +P+E + +P
 Sbjct: 799 YAPTEETTYAP-----TEETPYEPT-EETTYAPTEETPYEPTTEETTYPTTEETTYAPT 850

Query: 893 KEGLKYSFPGERPSHS 908
 +E Y+ P E+ +++
 Sbjct: 851 EE-TTYA-PTKTTYA 864

Score = 437 (65.6 bits), Expect = 3.3e-37, P = 3.3e-37
 Identities = 85/417 (20%), Positives = 223/417 (53%)

Query: 502 ERTPRGPSETRHNPSWRNHRSRSPERSQRSSLERRHHSRSPQRSHCSRSPKRNHSSPERSW 561
 E TP P+E T + P+ +P+E + + E+ ++P++ + +P+ + P+E +
 Sbjct: 419 EETPYEPTTEETTYPTTEETTYAPTEETTYAPTEKTTYAPTEETTYAPTEETPYEPTTEETT 478

Query: 562 RSPSRQNRHCSPPERSCHLSERGLHSPSRSHRGPSQRRHHSRSPERSHRSPERSHRSPS 621
 +P++ +P E + ++ +E ++P++ + P++ + P+E + +P+E + +P+
 Sbjct: 479 YAPTKETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPT 538

Query: 622 ERRHRSRSPQRSHRGPSERSHCSRSPERRHRSRSPQRSHRGPSERRHHSRSPKSRSHRSPARRSH 681
 E ++P++ + P+E + +P+E P++ + P+E ++P++ + +P +
 Sbjct: 539 EETTYAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETMYAPIEETT 598

Query: 682 RSPSERSHHSRSPERSHHSRSPERRHHSRSPERSHCSRSPERSHCSRSPERRHRSRSPERRHHSRSPS 741
 +P+E + ++P+E + + P+E ++P+E + +P+E + + +E +P+E ++P+
 Sbjct: 599 YAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYASTEETTYAPTEETTYAPA 658

Query: 742 EKSHHSRSPERSHHSRSPERRHHSPLERSRHSLLERSHRSPERRSHRSFERS-HRRISERS 800
 E++ + P+E + ++P+E ++P E + ++ E + +P+E ++ E + + E +
 Sbjct: 659 EETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTTEETT 718

Query: 801 HSPSEKSHLSPLERSRCSPEERRGHSSSGKTCHSPSERSHRSPSGMRQGRTERSERSHRSSC 860
 ++P+E++ +P E + +P E + + +T ++P+E + +P+ +E +
 Sbjct: 719 YAPTEETTYAPTEETMYAPIEETTYGPTTEETTYAPTEATTYAPTEETPYAPTEETTYEPT 778

Query: 861 ERTRHSPSEMRPGRPSGRNHCSPEERSRRSPLKEGLKYSFPGERPSHLSRDFKNQTT 918
 T ++P+E P+ +P+E + +P +E Y P E +++ + + T
 Sbjct: 779 GETTYAPTEETTYAPTEETTYAPTEETTYAPTEE-TPYE-PTEETTYAPTEETPYEPT 834

Score = 428 (64.2 bits), Expect = 3.1e-36, P = 3.1e-36
 Identities = 89/440 (20%), Positives = 228/440 (51%)

Query: 473 PGSPVKRTWHRHLKDKLTHKEHNHPSFYR-ERTPRGPSETRHNPSWRNHRSPSERSQRS 531
 P P + T + K+ T+ ++ E T P+E T + P+ P+E + +
 Sbjct: 470 PYEPTTEETTYAPTKET-TYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTTEETTYA 528

Query: 532 SLERRHHSQSRSHCSPSRKNHSSPSERSSWRSPSQRNHCSPPERSCHLSLSEGLHSPSQSR 591
 E ++P++ + +P+ + +P+E + +P++ P E + ++ +E ++P++
 Sbjct: 529 PTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEE 588

Query: 592 SHRGPSQRRHHSPEERSHRSPSERSHRSPERRHRSQSRSHRGPSERSHCSPEERRHRS 651
 + P + ++P+E + +P+E + P+E +P++ + P+E + + +E +
 Sbjct: 589 TMYAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYASTEETTYA 648

Query: 652 PSQSRSHRGPSERRHHSPEKSRSHSPARRSHSPERSHHSPEERSHHSPEERRHHSPEER 711
 P++ + P+E + P++ + +P + +P+E + ++P+E + ++P+E ++P+E
 Sbjct: 649 PTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAE 708

Query: 712 SHCSPERSHCSPEERRHRSPEERRHSPSEKSHHSPERSHHSPEERRHSPSERSRHS 771
 + P+E + +P+E +P+E ++P E++ + P+E + ++P+E ++P E + ++
 Sbjct: 709 TPYEPTTEETTYAPTEETTYAPTEETMYAPIEETTYGPTTEETTYAPTEATTYAPTEETPYA 768

Query: 772 LLERSHRSPERRSHRSFERS-HRRISERSHSPSEKSHLSPLERSRCSPEERRGHSSSGK 830
 E + P+ ++ E + + E ++P+E++ +P E + P+E ++ + +
 Sbjct: 769 PTEETTYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTTEETTYAPTEE 828

Query: 831 TCHSPSERSHRSPSGMRQGRTERSERSHRSSCERTRHSPSEMRPGRPSGRNHCSPEERSRRS 890
 T + P+E + +P+ +E + + E+T ++P+E P+ P+E + +
 Sbjct: 829 TPYEPTTEETTYPTTEETTYAPTEETTYAPTEKTTYAPTEETTYAPTEETPYEPTTEETTYA 888

Query: 891 PLKEGLKYSFPGERPSHLSRD 912
 P KE Y+ P E +++ + +
 Sbjct: 889 PTKE-TTYA-PTEETTYASTE 908

Score = 427 (64.1 bits), Expect = 4.0e-36, P = 4.0e-36
 Identities = 81/394 (20%), Positives = 213/394 (54%)

Query: 502 ERTPRGPSETRHNPSWRNHRSPSERSQRSLSLERRHHSQSRSHCSPSRKNHSSPSERSW 561
 E T GP+E T + P+ +P+E + + E + P+ + +P+ + +P+E +
 Sbjct: 739 EETTYGPTTEETTYAPTEATTYAPTEETPYAPTEETTYEPTGETTYAPTEETTYAPTEETT 798

Query: 562 RSPSQRNHCSPPERSCHLSLSEGLHSPSQSRSHRGPSQRRHHSPEERSHRSPSERSHRSPS 621
 +P++ +P E + + +E ++P++ + P++ ++P+E + +P+E + +P+
 Sbjct: 799 YAPTEETTYAPTEETPYEPTTEETTYAPTEETPYEPTTEETTYPTTEETTYAPTEETTYAPT 858

Query: 622 ERRHRSQSRSHRGPSERSHCSPEERRHRSQSRSHRGPSERRHHSPEKSRSHSPARRSH 681
 E+ +P++ + P+E + P+E +P++ + P+E ++ ++ + +P +
 Sbjct: 859 EKTYYAPTEETTYAPTEETPYEPTTEETTYAPTEETTYAPTEETTYASTEETTYAPTEETT 918

Query: 682 RSPERSHHSPEERSHHSPEERRHHSPEERSHCSPEERSHCSPEERRHRSPEERRHSPS 741
 +P+E + + P+E + ++P+E ++P+E + +P+E + +P+E +P+E + P+
 Sbjct: 919 YAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPT 978

Query: 742 EKSHHSPERSHHSPEERRHSPSERSRHSLLERSHRSPERRSHRSFERS-HRRISERS 800
 E++ ++P+E + ++P+E ++P+E + ++ E + +P+E + E + + E +
 Sbjct: 979 EETTYAPTEETTYAPTEETMYAPIEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETT 1038

Query: 801 HSPSEKSHLSPLERSRCSPEERRGHSSSGKTCHSPSERSHRSPSGMRQGRTERSERSHRSSC 860
 ++P+E++ + E + +P+E ++ + +T + P+E + +P+ +E + +
 Sbjct: 1039 YAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPT 1098

Query: 861 ERTRHSPSEMRPGRPSGRNHCSPEERSRRSPLKE 894
 E T ++P+E P+ P+E + +P +E
 Sbjct: 1099 EETTYAPTEETTYAPAEETPYEPTTEETTYAPTEE 1132

Score = 424 (63.6 bits), Expect = 8.5e-36, P = 8.5e-36
 Identities = 81/394 (20%), Positives = 210/394 (53%)

Query: 502 ERTPRGPSETRHNPSWRNHRSPSERSQRSLSLERRHHSQSRSHCSPSRKNHSSPSERSW 561
 E T P+E T + P+ +P+E + + E + P++ + +P+ + +P+E +
 Sbjct: 939 EETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETM 998

Query: 562 RSPSQRNHCSPERSCHSLSERGLHSPSQSRHSGPSQRRHSPSERSHRSPSERSHRSPS 621
 +P + +P E + ++ +E + P++ + P++ ++P+E + + +E + +P+
 Sbjct: 999 YAPIEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYASTEETTYAPT 1058

Query: 622 ERRHRSQSRHSGPSERSHSCSPSERRRHRSQSRHSGPSERRHHSKRSRHRSPARRSH 681
 E +P++ + P+E + +P+E +P++ + P+E ++P++ + +PA +
 Sbjct: 1059 EETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETP 1118

Query: 682 RSPSERSHHSPSERSHHSPERRHHSERSHSCSPSERSHSCSPERRHRSPPERRHHS 741
 P+E + ++P+E + ++P+E ++P E + P+E + +P+E +P+E ++P+
 Sbjct: 1119 YEPTTEETTYAPTEETTYAPTEETMYAPIEETTYGPTTEETTYAPTEATTYAPTEETPYAPT 1178

Query: 742 EKSHHSPSERSHHSPERRHSPERSRHSLLERSHRSPSERRSHRSFERS-HRRISERS 800
 E++ + P+ + ++P+E ++P E + ++ E + +P+E + E + + E +
 Sbjct: 1179 EETTYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTTEETTYAPTEETT 1238

Query: 801 HSPSEKSHLSPLERSRCSPPSERRGHSSSGKTCHSPSERSHRSPSGMRQGRTERSRRSSC 860
 + P+E++ +P E + +P+E ++ + +T ++P + + P+ +E + +
 Sbjct: 1239 YEPTTEETTYAPTEETTYAPTEETTYAPTEETMYAPIDETYYGPTTEETTYAPTEATTYAPT 1298

Query: 861 ERTRHSPSEMRRPGRPSGRNHCSPPSERRRSPKE 894
 E T ++P+E P+G +P+E + +P +E
 Sbjct: 1299 EETPYAPTEETTYEPTGETTYAPTEETTYAPTEE 1332

Score = 422 (63.3 bits), Expect = 1.4e-35, P = 1.4e-35
 Identities = 84/407 (20%), Positives = 216/407 (53%)

Query: 502 ETRPRGSPSETRHNPSPWRNHRSPPSERSQSRSSLERHHSPSQSRHSCSPSRKNHSSPERSW 561
 E T P+E T + P+ P+E + + E + P++ + +P+ +P+E +
 Sbjct: 795 EETTYAPTEETTYAPTEETPYEPTTEETTYAPTEETPYEPTTEETTYPTTEETTYAPTEETT 854

Query: 562 RSPSQRNHCSPERSCHSLSERGLHSPSQSRHSGPSQRRHSPSERSHRSPSERSHRSPS 621
 +P+++ +P E + ++ +E + P++ + P++ ++P+E + + +E + +P+
 Sbjct: 855 YAPTEKTTYAPTEETTYAPTEETPYEPTTEETTYAPTKETTYAPTEETTYASTEETTYAPT 914

Query: 622 ERRHRSQSRHSGPSERSHSCSPSERRRHRSQSRHSGPSERRHHSKRSRHRSPARRSH 681
 E +P++ + P+E + +P+E +P++ + P+E ++P++ + +PA +
 Sbjct: 915 EETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETP 974

Query: 682 RSPSERSHHSPSERSHHSPERRHHSERSHSCSPSERSHSCSPERRHRSPPERRHHS 741
 P+E + ++P+E + ++P+E ++P E + +P+E + +P+E P+E ++P+
 Sbjct: 975 YEPTTEETTYAPTEETTYAPTEETMYAPIEETTYAPTEETTYAPAEETPYEPTTEETTYAPT 1034

Query: 742 EKSHHSPSERSHHSPERRHSPERSRHSLLERSHRSPSERRSHRSFERS-HRRISERS 800
 E++ ++P+E + ++ +E ++P E + ++ E + P+E ++ E + + E +
 Sbjct: 1035 EETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETT 1094

Query: 801 HSPSEKSHLSPLERSRCSPPSERRGHSSSGKTCHSPSERSHRSPSGMRQGRTERSRRSSC 860
 ++P+E++ +P E + +P+E + + +T ++P+E + +P+ E +
 Sbjct: 1095 YAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETMYAPIEETTYGPT 1154

Query: 861 ERTRHSPSEMRRPGRPSGRNHCSPPSERRRSPKEGLKYSFPGERPSHS 908
 E T ++P+E P+ +P+E + P E Y+ P E +++
 Sbjct: 1155 EETTYAPTEATTYAPTEETPYAPTEETTYEPTGE-TTYA-PTEETTYA 1200

Score = 421 (63.2 bits), Expect = 1.8e-35, P = 1.8e-35
 Identities = 86/418 (20%), Positives = 219/418 (52%)

Query: 491 HKEHNHPSFYRETRPRGSPSETRHNPSPWRNHRSPPSERSQSRSSLERHHSPSQSRHSCSPSR 550
 H H E T P+E T + P+ +P+E + + E + P++ + +P+
 Sbjct: 376 HYAHIEKPCDTEVTMYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTTEETTYPTTE 435

Query: 551 KNHSSPERSWRSPSQRNHCSPERSCHSLSERGLHSPSQSRHSGPSQRRHSPSERSHR 610
 + +P+E + +P+++ +P E + ++ +E + P++ + P++ ++P+E +
 Sbjct: 436 ETTYAPTEETTYAPTEKTTYAPTEETTYAPTEETPYEPTTEETTYAPTKETTYAPTEETTY 495

Query: 611 SPERSHRSPSERRHRSQSRHSGPSERSHSCSPSERRRHRSQSRHSGPSERRHHSKRSRHRSP 670
 + +E + +P+E +P++ + P+E + +P+E +P++ + P+E ++P++
 Sbjct: 496 ASTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE 555

Query: 671 RSHRSPARRSHRSPPSERSHHSPSERSHHSPERRHHSERSHSCSPSERSHSCSPERRHR 730
 + +PA + P+E + ++P+E + ++P+E ++P E + +P+E + +P+E
 Sbjct: 556 ETTYPAEETPYEPTTEETTYAPTEETTYAPTEETMYAPIEETTYAPTEETTYAPAEETPY 615

Query: 731 SPERRHHSPEKSHHSPSERSHHSPERRHSPERSRHSLLERSHRSPSERRSHRSFE 790
 P+E ++P+E++ ++P+E + ++ +E ++P E + ++ E + P+E ++ E
 Sbjct: 616 EPTTEETTYAPTEETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTE 675

Query: 791 RS-HRRISERSHSPSEKSHLSPLERSRCSPPSERRGHSSSGKTCHSPSERSHRSPSGMRQG 849
 + + E +++P+E++ +P E + +P+E + + +T ++P+E + +P+

Sbjct: 676 ETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETMY 735

Query: 850 RTSERSHRSSCERTRHSPSEMRPGRPSGRNHCSPSERSRRSPLKEGLKYSPFGERPSHS 908
E T ++P+E P+ +P+E + P E Y+ P E +++

Sbjct: 736 APIEETTYGPTTEETTYAPTEATTYAPTEETPYAPTEETTYEPTGE-TTYA-PTEETTYA 792

Score = 420 (63.0 bits), Expect = 2.3e-35, P = 2.3e-35
Identities = 82/393 (20%), Positives = 206/393 (52%)

Query: 502 ERTPRGFSERTRHNPSWRNHRSPSERSQSSSLERRHHSQSRSHCSPSRKNHSSPSERSW 561
E TP P+E T + P+ +P+E + + +E ++P++ + +P+ + P+E +

Sbjct: 971 EETPYEPTTEETTYAPTEETTYAPTEETMYAPIEETTYAPTEETTYAPAEETPYEPTTEET 1030

Query: 562 RSPSQRNHCSPERSCHLSERGLHSPSQSRHGRPSQRRHHSPSERSHRSPERSHRSPS 621
+P++ +P E + ++ +E ++P++ + P++ + P+E + +P+E + +P+

Sbjct: 1031 YAPTEETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPT 1090

Query: 622 ERRHRSQSRSHRGPSERSHCSPSERRHRSQSRSHRGPSERRHHSPSKRSHRSPARRSH 681
E +P++ + P+E + +P+E P++ + P+E ++P++ + +P +

Sbjct: 1091 EETTYAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETMYAPIEETT 1150

Query: 682 RSPSERSHHSPERSHHSPERRHHSPSERSHCSPSERSHCSPSERRHRSPSERRHHS 741
P+E + ++P+E + ++P+E ++P+E + P+ + +P+E +P+E ++P+

Sbjct: 1151 YGPTTEETTYAPTEATTYAPTEETPYAPTEETTYEPTGETTYAPTEETTYAPTEETTYAPT 1210

Query: 742 EKSHHSPERSHHSPERRHHSPLERSRHSLLERSHRSPERRSHRSFERS-HRRISERS 800
E++ ++P+E + + P+E ++P E + + E + +P+E ++ E + + E

Sbjct: 1211 EETTYAPTEETPYEPTTEETTYAPTEETTYEPTTEETTYAPTEETTYAPTEETTYAPTEETM 1270

Query: 801 HSPSEKSHLSPLERSRCSPSERRGHSSGKTCHSPERSHRSPSGMRQGRTERSRRSSC 860
++P +++ P E + +P+E ++ + +T ++P+E + P+G +E + +

Sbjct: 1271 YAPIDETTYGPTTEETTYAPTEATTYAPTEETPYAPTEETTYEPTGETTYAPTEETTYAPT 1330

Query: 861 ETRHSPSEMRPGRP-----SGRNHCSPSE 885

Sbjct: 1331 EETTYAPMEETPYEPAEESTSTVSTKPCNTEE 1363

Score = 419 (62.9 bits), Expect = 3.0e-35, P = 3.0e-35
Identities = 83/411 (20%), Positives = 215/411 (52%)

Query: 502 ERTPRGFSERTRHNPSWRNHRSPSERSQSSSLERRHHSQSRSHCSPSRKNHSSPSERSW 561
E T P+E T + P+ +P+E + E ++P++ + +P+ + +P E +

Sbjct: 947 EETTYAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETMYAPIEETT 1006

Query: 562 RSPSQRNHCSPERSCHLSERGLHSPSQSRHGRPSQRRHHSPSERSHRSPERSHRSPS 621
+P++ +P E + + +E ++P++ + P++ ++ +E + +P+E + +P+

Sbjct: 1007 YAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYASTEETTYAPTEETTYAPA 1066

Query: 622 ERRHRSQSRSHRGPSERSHCSPSERRHRSQSRSHRGPSERRHHSPSKRSHRSPARRSH 681
E P++ + P+E + +P+E +P++ + P+E ++P++ + P +

Sbjct: 1067 EETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTTEETT 1126

Query: 682 RSPSERSHHSPERSHHSPERRHHSPSERSHCSPSERSHCSPSERRHRSPSERRHHS 741
+P+E + ++P+E + ++P E + P+E + +P+E + +P+E +P+E + P+

Sbjct: 1127 YAPTEETTYAPTEETMYAPIEETTYGPTTEETTYAPTEATTYAPTEETPYAPTEETTYEPT 1186

Query: 742 EKSHHSPERSHHSPERRHHSPLERSRHSLLERSHRSPERRSHRSFERS-HRRISERS 800
++ ++P+E + ++P+E ++P E + ++ E + P+E ++ E + + E +

Sbjct: 1187 GETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTTEETTYAPTEETTYEPTTEETT 1246

Query: 801 HSPSEKSHLSPLERSRCSPSERRGHSSGKTCHSPERSHRSPSGMRQGRTERSRRSSC 860
++P+E++ +P E + +P+E ++ +T + P+E + +P+ +E + +

Sbjct: 1247 YAPTEETTYAPTEETTYAPTEETMYAPIDETTYGPTTEETTYAPTEATTYAPTEETPYAPT 1306

Query: 861 ETRHSPSEMRPGRPSGRNHCSPSERSRRSPLKEGLKYSPFGERPSHSLSRD 912

Sbjct: 1307 EETTYEPTGETTYAPTEETTYAPTEETTYAPMEE-TPYE-PAEESTSTVSTE 1356

Score = 415 (62.3 bits), Expect = 8.0e-35, P = 8.0e-35
Identities = 84/423 (19%), Positives = 218/423 (51%)

Query: 473 PGSPVKRTWHRHLKDKLTHKEHNHPSFYR-ERTPRGFSERTRHNPSWRNHRSPSERSQRS 531
P P + T + K+ T+ ++ E T P+E T + P+ P+E + +

Sbjct: 878 PYEPTTEETTYAPTKET-TYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTTEETTYA 936

Query: 532 SLERRHHSQSRSHCSPSRKNHSSPSERSWRSPSQRNHCSPERSCHLSERGLHSPSQSR 591
E ++P++ + +P+ + +P+E + +P++ P E + ++ +E ++P++

Sbjct: 937 PTEETTYAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEE 996

Query: 592 SHRGPSQRRHHSPSERSHRSPERSHRSPERRHRSQSRSHRGPSERSHCSPSERRHRS 651
+ P + ++P+E + +P+E + P+E +P++ + P+E + + +E +

Sbjct: 997 TMYAPIEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYASTEETTYA 1056

Query: 652 PSQRSHRGPSERRHHSKRSRSHRSPARRSHRSPSERSHHSPSERSHHSPSERRHHSPSER 711
 P++ + P+E + P++ + +P + +P+E + ++P+E + ++P+E ++P+E

Sbjct: 1057 PTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAE 1116

Query: 712 SHCSPSERSHCSPSERRHRSFERS-HRRISERSHSPSEKSHLSPLERSRCSPSERRHSSSGK 771
 + P+E + +P+E +P+E ++P E++ + P+E + ++P+E ++P E + ++

Sbjct: 1117 TPYEPTTEETTYAPTEETTYAPTEETMYAPIEETTYGPTTEETTYAPTEATTYAPTEETPYA 1176

Query: 772 LLSRSHRSPSERRSHRSFERS-HRRISERSHSPSEKSHLSPLERSRCSPSERRHSSSGK 830
 E + P+ ++ E + + E +++P+E++ +P E + P+E ++ + +

Sbjct: 1177 PTEETTYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTTEETTYAPTEE 1236

Query: 831 TCHSPSERSHRSPSGMRQRTSERSHRSSCERTRHSPSEMGRPGRNHCSPSERSRRS 890
 T + P+E + +P+ +E + + E T ++P + P+ +P+E + +

Sbjct: 1237 TTYEPTTEETTYAPTEETTYAPTEETTYAPTEETMYAPIDETYYGPTTEETTYAPTEATTYA 1296

Query: 891 PLKE 894
 P +E

Sbjct: 1297 PTEE 1300

Score = 403 (60.5 bits), Expect = 1.6e-33, P = 1.6e-33
 Identities = 84/394 (21%), Positives = 213/394 (54%)

Query: 501 RERTPRGSPSERTRHNPSWRNHRSPSERSQRSSLERHHSPSQRSHCSPSRKNHSSPSERS 560
 RE T PSE T + P +P+E+ +E + + ++ +P++ ++P+ER

Sbjct: 319 REETAAPSEDTTYAPREVTPYAPTEKPY--DVEETTYVTEESTY-APTKSETNAPTERM 375

Query: 561 WRSPQRNHCSPPERSCHLSERGLHSPSQRSHRGPSQRHHSPSERSHRSPSERSHRSP 620
 + ++ C E + ++ +E ++P++ + P++ ++P+E + P+E + +P

Sbjct: 376 HYAHIEKP-CDT-EVTMYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTTEETTYTP 433

Query: 621 SERRHRSPSQRSHRGPSERSHCSPSERRHRSFERS-HRRISERSHSPSEKSHLSPLERSRCSPSERRHSSSGK 680
 +E +P++ + P+E++ +P+E +P++ + P+E ++P+K + +P +

Sbjct: 434 TEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTTEETTYAPTEETTYAPTEET 493

Query: 681 HRSPSERSHHSPSERSHHSPSERRHHSPSERSHCSPSERSHCSPSERRHRSFERS-HRRISER 740
 + +E + ++P+E + ++P+E + P+E + +P+E + +P+E +P+E ++P

Sbjct: 494 TYASTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAP 553

Query: 741 SEKSHHSPSERSHHSPSERRHHSPLERSRHSLLERSHRSPSERRSHRSFERS-HRRISER 799
 +E++ ++P+E + + P+E ++P E + ++ E + +P E ++ E + + E

Sbjct: 554 TEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETMYAPIEETTYAPTEETTYAPAEET 613

Query: 800 SHSPSEKSHLSPLERSRCSPSERRHSSSGKTCHSPSERSHRSPSGMRQRTSERSHRSS 859
 + P+E++ +P E + +P+E ++S+ +T ++P+E + +P+ +E + +

Sbjct: 614 PYEPTTEETTYAPTEETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTTEETTYAP 673

Query: 860 CERTRHSPSEMGRPGRNHCSPSERSRRSPLKE 894
 E T ++P+E P+ +P+E + +P +E

Sbjct: 674 TEETTYAPTEETTYAPTEETTYAPTEETTYAPAE 708

Score = 398 (59.7 bits), Expect = 5.5e-33, P = 5.5e-33
 Identities = 84/402 (20%), Positives = 209/402 (51%)

Query: 475 SPVKRTWHRHLKDLTHKEHNHPSFY-RERTPRGSPSERTRHNPSWRNHRSPSERSQRSSL 533
 +P + T + +++ T+ ++ E TP P+E T + P+ +P+E + +S

Sbjct: 992 APTEETMYAPIEET-TYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAST 1050

Query: 534 ERRHHSQSPSQRSHCSPSRKNHSSPSERSWRSQSPNHCSPERSCHLSERGLHSPSQRSH 593
 E ++P++ + +P+ + P+E + +P++ +P E + ++ +E ++P++ +

Sbjct: 1051 EETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETT 1110

Query: 594 RGPSQRHHSPSERSHRSPSERSHRSPSERRHRSFERS-HRRISERSHSPSEKSHLSPLERSRCSPSERRHSSSGK 653
 P++ + P+E + +P+E + +P+E +P + + GP+E + +P+E +P+

Sbjct: 1111 YAPAEETPYEPTTEETTYAPTEETTYAPTEETMYAPIEETTYGPTTEETTYAPTEATTYAPT 1170

Query: 654 QRSHRGPSERRHHSKRSRSHRSPARRSHRSPSERSHHSPSERSHHSPSERRHHSPSERSH 713
 + + P+E + P+ + +P + +P+E + ++P+E + ++P+E + P+E +

Sbjct: 1171 EETPYAPTEETTYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTTEETT 1230

Query: 714 CSPSERSHCSPSERRHRSFERS-HRRISERSHSPSEKSHLSPLERSRCSPSERRHSSSGKTCH 773
 +P+E + P+E +P+E ++P+E++ ++P+E + ++P + + P E + ++

Sbjct: 1231 YAPTEETTYEPTTEETTYAPTEETTYAPTEETTYAPTEETMYAPIDETYYGPTTEETTYAPT 1290

Query: 774 ERSRSPSERRSHRSFERS-HRRISERSHSPSEKSHLSPLERSRCSPSERRHSSSGKTCH 833
 E + +P+E + E E ++ P+ ++ +P E + +P+E ++ +T +

Sbjct: 1291 EATTYAPTEETPYAPTE-----ETTYEPTGETTYAPTEETTYAPTEETTYAPMEETPY 1343

Query: 834 SPERSHRSPSGMRQRTSERSHRSSCERTRHSPSEMGRPGRS 876

P+E S + S + T E + + E T PS+ P+
 Sbjct: 1344 EPAEESTSTVSTKPCNTEETDEPTDEPT-DEPSDEPTDEPT 1385

Score = 368 (55.2 bits), Expect = 9.5e-30, P = 9.5e-30
 Identities = 79/386 (20%), Positives = 211/386 (54%)

Query: 524 PSERSQRSSLERRHHSQRSRSHCSPSRKNHSSPSERSWRSPSQRNHCSPPERSCHSLSER 583
 PS+ ++ + E + P + + +PS +P E + +P+++ + E + + ++E

Sbjct: 303 PSDTEAPT-EGTTYVPREETTAAPSEDTTYAPREVTYPAPTEKPY--DVEETTY-VTEE 358

Query: 584 GLHSPQRSRSHRGPSQRRHHSRSPSER-----SHRSPSERSHRSPSERRHRSQRSRSHRGPS 637
 ++P++ P++R H++ E+ + +P+E + +P+E +P++ + P+

Sbjct: 359 STYAPTKESETNAPTERMHYAHIEKPCDTEVTMYAPTEETTYAPTEETTYAPTEETTYAPT 418

Query: 638 ERSRSHCSPPERSHRSPQRSRSHRGPSERRHHSRSHRSPARRSHRSPSERSHHSRSPSERH 697
 E + P+E +P++ + P+E ++P++++ +P + +P+E + + P+E +

Sbjct: 419 EETPYEPTETTYTPTETTYAPTEETTYAPTEKTTYAPTEETTYAPTEETPYEPTETTT 478

Query: 698 HSPSERRHHSRSPERSHCSPPERSHCSPPERRHRSRSPERRHHSRSPSEKSHHSRSPERSHHS 757
 ++P++ ++P+E + + +E + +P+E +P+E + P+E++ ++P+E + ++P+

Sbjct: 479 YAPTKETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTETTYAPTEETTYAPT 538

Query: 758 ERRRHSPLERSRHSLLERSHRSPERRSHRSFERS-HRRISERSHSPSEKSHLSPLERSR 816
 E ++P E + ++ E + +P+E + E + + E +++P+E++ +P+E +

Sbjct: 539 EETTYAPTEETTYAPTEETTYAPAEETPYEPTETTYAPTEETTYAPTEETMYAPIEETT 598

Query: 817 CSPSERRGHSSSGKTCRSPSERSHRSPSGMRQGRTERSRRSCERTRHSRSPSEMRPGRPS 876
 +P+E ++ + +T + P+E + +P+ +E + +S E T ++P+E P+

Sbjct: 599 YAPTEETTYAPAEETPYEPTETTYAPTEETTYAPTEETTYASTEETTYAPTEETTYAPA 658

Query: 877 GRNHCSPPERSRRSPLKEGLKYSFPGERPSSHS 908
 P+E + +P +E Y+ P E +++

Sbjct: 659 EETPYEPTETTYAPTEE-TTYA-PTEETTYA 688

Score = 337 (50.6 bits), Expect = 2.1e-26, P = 2.1e-26
 Identities = 66/328 (20%), Positives = 170/328 (51%)

Query: 502 ERTPRGSPERTRHNPWRNHRSPSERSQRSSLERRHHSQRSRSHCSPSRKNHSSPSERSW 561
 E T P+E T + P+ +P+E + + E ++P++ + +P+ + +P+E +

Sbjct: 1059 EETTYAPAEETPYEPTETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETP 1118

Query: 562 RSPSQRNHCSPPERSCHSLSERGLHSPQRSRSHRGPSQRRHHSRSPSERSHRSPSERSHRSPS 621
 P++ +P E + ++ +E +++P + + GP++ ++P+E + +P+E + +P+

Sbjct: 1119 YEPTETTYAPTEETTYAPTEETMYAPIEETTYGPTETTYAPTEATTYAPTEETPYAPT 1178

Query: 622 ERRHRSQRSRSHRGPSERSHCSPPERRHRSQRSRSHRGPSERRHHSRSPKRSHRSPARRSH 681
 E P+ + P+E + +P+E +P++ + P+E + P++ + +P +

Sbjct: 1179 EETTYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTETTYAPTEETT 1238

Query: 682 RSPSERSHHSRSPERSHHSRSPERRHHSRSPERSHCSPPERSHCSPPERRHRSRSPERRHHS 741
 P+E + ++P+E + ++P+E ++P+E + +P + + P+E +P+E ++P+

Sbjct: 1239 YEPTETTYAPTEETTYAPTEETTYAPTEETMYAPIDETYYGPTETTYAPTEATTYAPT 1298

Query: 742 EKSHHSPSERSHHSRSPERRHHSPLERSRHSLLERSHRSPERRSHRSFERSHRRIS---- 797
 E++ ++P+E + + P+ ++P E + ++ E + +P E + E S +S

Sbjct: 1299 EETPYAPTEETTYEPTGETTYAPTEETTYAPTEETTYAPMEETPYEPAEESTSTVSTKPC 1358

Query: 798 ----ERSHSPSEKSHLSPLERSRCSPE 821
 E + P+++ P + P++

Sbjct: 1359 CNTEETDEPTDEPTDEPSDEPTDEPTD 1386

Score = 333 (50.0 bits), Expect = 5.7e-26, P = 5.7e-26
 Identities = 63/320 (19%), Positives = 166/320 (51%)

Query: 502 ERTPRGSPERTRHNPWRNHRSPSERSQRSSLERRHHSQRSRSHCSPSRKNHSSPSERSW 561
 E T P+E T + P+ +P+E + + E ++P++ + P+ + +P+E +

Sbjct: 1075 EETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTETTYAPTEETT 1134

Query: 562 RSPSQRNHCSPPERSCHSLSERGLHSPQRSRSHRGPSQRRHHSRSPSERSHRSPSERSHRSPS 621
 +P++ +P E + + +E ++P++ + P++ ++P+E + P+ + +P+

Sbjct: 1135 YAPTEETMYAPIEETTYGPTETTYAPTEATTYAPTEETPYAPTEETTYEPTGETTYAPT 1194

Query: 622 ERRHRSQRSRSHRGPSERSHCSPPERRHRSQRSRSHRGPSERRHHSRSPKRSHRSPARRSH 681
 E +P++ + P+E + +P+E P++ + P+E + P++ + +P +

Sbjct: 1195 EETTYAPTEETTYAPTEETTYAPTEETPYEPTETTYAPTEETTYEPTETTYAPTEETT 1254

Query: 682 RSPSERSHHSRSPERSHHSRSPERRHHSRSPERSHCSPPERSHCSPPERRHRSRSPERRHHS 741
 +P+E + ++P+E + ++P + + P+E + +P+E + +P+E +P+E + P+

Sbjct: 1255 YAPTEETTYAPTEETMYAPIDETYYGPTETTYAPTEATTYAPTEETPYAPTEETTYEPT 1314

Query: 742 EKSHHSPSERSHHSRSPERRHHSPLERSRHSLLERSHRSPERRSHRSFERSHRRISERSH 801

Score = 303 (45.5 bits), Expect = 9.6e-23, P = 9.6e-23
Identities = 70/322 (21%), Positives = 170/322 (52%)

```
Query:      883 PSERSRRSPLKEGLKYSFPGERP 905
              P E + +P +E Y+ E P
Sbjct:     593 PIEETTYAPTEE-TTYAPAEETP 614
```

```
Query:      889 RSPLKEGLKYSFPGERPSHLSRD 912
              +P KE   Y+ P E   +++ + +
Sbjct:      479 YAPTKE-TTYA-PTEETTYASTEE 500
```

Report for DKFZphtes3 8q11.2

```
SEQ      ESSLSIFYDREDLVPMEESDSQSDSQTRISESHSLKPNYLQAQTDFSEQQLLEDLQ
SEG      .....XXXXXXXXXXXX.....
PRD      cccceccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhh

SEQ      LKIAAKLLRSQIPDPVPPPLASGLVKYPICLCGRCSGLNCHHKLQTTSGPYLLIYPQL
SEG      .....
PRD      hhhhhhhhhhhccccccccccccceeeecceecccccccccccccccccceeeehhh

SEQ      HLVRTPEGHGEVRLHLGFLRLIGKRSQISKYRERDRPVIIRSPISPSORKAKIYTQASKS
SEG      .....
PRD      hccccccccceeecccceccccccccccccccccceeeecccccchhhhhhhcccc

SEQ      PTSTIDLQSGPSQPAPVQVYIRRQRSRPDLVEKTKTRAPHGYEFTQVHNLPESDSEST
```



```
SEG .....
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccch

SEQ QNEKRAKVRTKKTSDSKYPMKRITKRLRKHRYFTNSRTTIESPSRELAHLRRKRIGAT
SEG .....
PRD hhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ QTSTASLKRQPKKPSQPKFMQLLFQSLKRAFQTAHRVIASVGRKPDGTRPDNLWASKNY
SEG .....
PRD cccchhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ YPKQNARDYCLPSSIKRDKRSADKLT PAGSTIKQEDILWGGTVQCRSAQQPRRAYSFQPR
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ PLRLPKPTDSQSGIAFQTASVGQPLRTVQKDSRSSRSKKNFYRNETSSQESKNLSTPGTRV
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ QARGRILPGSPVKRTWHRHLKDKLTHKEHNHPSFYRERTPRGPSETRHNPSWRNHRSPS
SEG .....
PRD eeeeecccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ ERSQRSSLERRHHSQSRSHCSPSRKNNHSSPSERSWRSPSQRNHCSPPERSCHLSERGL
SEG .....
PRD chhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ HSPSQSRSHRGPSQRRHHSPSERSHRSPERSHRSPERRHRSQSRSHRGPSERSHCSPS
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ ERRHRSQSRSHRGPSERRHHSPSKRSRSPARRSHRSPSERSHHSPSERSHHSPSERRH
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ HSPERSHCSPSERSHCSPSERRHRSPSERRHHSPEKSHHSPSERSHHSPSERRHRSPL
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ ERSRHSLLERSHRSPERRSHRSFERSHRRISERSHSPSEKSHLSPLERSRCSPSERRGH
SEG .....
PRD hhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ SSSGKTCHSPERSHRSPSGMRQGTSESRSHRSSCERTRHSPSEMRPGRPSGRNHCSPE
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ RSRRSPLKEGLKYSFPGERPSHSLSRDFKNQTTLLGTTKNPKAGQVWRPEATR
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
```

Prosites for DKFZphtes3_8g11.2

PS00017 839->847 ATP_GTP_A PDOC00017

(No Pfam data available for DKFZphtes3_8g11.2)

DKFZphtes3_8g5

group: testes derived

DKFZphtes3_8g5 encodes a novel 544 amino acid protein nearly identical to human KIAA087 protein.

The novel protein is a new splice variant of KIAA087.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

KIAA087, alternative spliced

complete cDNA, complete cds, EST hits

Sequenced by MediGenomix

Locus: unknown

Insert length: 2762 bp

No poly A stretch found, no polyadenylation signal found

```

1 CCGACATCGG CCGTGCTCTCC AGCACCTGCC GCGGGCTGCG CGAGCTGTGC
51 CAGAGCAGCG GGAAGGTGTG GAAGGAGCAG TTCCGGGTGA GGTGACCTTC
101 CCTTATGAAA CACTACAGCC CCACCGACTA CGTCAATTGG TTGGAAGAGT
151 ATAAAGTTTCG GCAAAAAGCT GGGTTAGAAG CGCGGAAGAT TGTAGCCTCG
201 TTCTCAAAGA GGTTCCTTTC AGAGCACGTT CCTTGTAATG GCTTCAGTGA
251 CATTGAGAAC CTTGAAGGAC CAGAGATTTT TTTTGAGGAT GAAGTGGTGT
301 GTATCCTAAA TATGGAAGGA AGAAAAGCTT TGACCTGGAA ATACTACGCA
351 AAAAAAATTC TTTACTACCT GCGGCAACAG AAGATCTTAA ATAATCTTAA
401 GGCCTTTCTT CAGCAGCCAG ATGACTATGA GTCGTATCTT GAAGGTGCTG
451 TATATATTGA CCAGTACTGC AATCCTCTCT CCGACATCAG CCTCAAAGAC
501 ATCCAGGCCC AAATTGACAG CATCGTGGAG CTGTTTGGCA AAACCTTTCG
551 GGGCATAAAC AGTCGCCACC CCAGCTTGGC CTTCAAGGCA GGTGAATCAT
601 CCATGATAAT GGAAATAGAA CTCCAGAGCC AGGTGCTGGA TGCCATGAAC
651 TATGTCCTTT ACGACCAACT GAAGTTCAAG GGAATCGAA TGGATTACTA
701 TAATGCCCTC AACTTATATA TGCATCAGGT TTTGATTCGC AGAACAGGAA
751 TCCCAATCAG CATGTCTCTG CTCTATTGGA CAATTGCTCG GCAGTTGGGA
801 GTCCCACTGG AGCCTGTCAA CTCCCAAGT CACTTCTTAT TAAGTGGTGT
851 CCAAGGCGCA GAAGGGGCGA CCCTGGACAT CTTTGACTAC ATCTACATAG
901 ATGCTTTTGG GAAAGGCAAG CAGCTGACAG TGAAGAATG CGAGTACTTG
951 ATCGGCCAGC ACGTGACTGC AGCACTGTAT GGGGTGGTCA ATGTCAAGAA
1001 GGTGTTACAG AGAATGGTGG GAAACCTGTT AAGCCTGGGG AAGCGGGAAG
1051 GCATCGACCA GTCATACCAG CTCCTGAGAG ACTCGCTGGA TCTCTATCTG
1101 GCAATGTACC CGGACCAGGT GCAGCTTCTC CTCCTCCAAG CCAGGCTTTA
1151 CTTCCACCTG GGAATCTGGC CAGAGAAGTC TTTCTGTCTT GTTTTGAAGG
1201 TGCTTGACAT CCTCCAGCAC ATCCAAACCC TAGACCCGGG GCAGCACGGG
1251 GCGGTGGGCT ACCTGGTGCA GCACACTCTA GAGCACATTG AGCGCAAAAA
1301 GGAGGAGGTG GCGGTAGAGG TGAAGCTGCG CTCCGATGAG AAGCACAGAG
1351 ATGCTGTCTA CTCCATCGGG CTCATTATGA AGCATAAGAG GTATGGCTAT
1401 AACTGTGTGA TCTACGGCTG GGACCCACCC TGCATGATGG GACACGAGTG
1451 GATCCGGAAC ATGAACGTCC ACAGCCTGCC GCACGGCCAC CACCAACCTT
1501 TCTATAACGT GCTGCTGGAG GACGGCTCCT GTCGATACGC AGCCCAAGAA
1551 AACTTGGAAT ATAACGTGGA GCCTCAAGAA ATCTCACACC CTGACGTGGG
1601 ACGCTATTTC TCAGAGTTTA CTGGCACTCA CTACATCCCA AACGCAGAGC
1651 TGGAGATCCG GTATCCAGAA GATCTGGAGT TTGTCTATGA AACGGTGCAG
1701 AATATTTACA GTGCAAGAAA AGAGAACATA GATGAGTAAA GTCTAGAGAG
1751 GACATTGCAC CTTTGCTGCT GCTGCTATCT TCCAAGAGAA CGGGACTCCG
1801 GAAGAAGACG TCTCCACGGA GCCCTCGGGA CCTGCTGCAC CAGGAAAGCC
1851 ACTCCACCAG TAGTGCTGGT TGCCTCCTAC TAACTTTAAA TACCGTGTGC
1901 TCTTCCCCAG CTGCAAGAGC AATGTTGCTC TCCGCTTACA CTAGTGAATT
1951 AATCTGAAAG GCACTGTGTC AGTGGCATGG CTGTATGCTT TGTCTGTGG
2001 TGACAGTTTG TGACATTCTG TCTTCATGAG GTCTCACAGT CGACGCTCCT
2051 GTAATCATTC TTTGTATTCA CTCCATTCCC CTGTCTGTCT GCATTGTGCT
2101 CAGAACATTT CCTTGGCTGG ACAGATGGGG TTAGTCATTT GCAATAATTT
2151 CCTTCTGATT TCTCTGTGGA ACGTGTTCGG TCCCGAGTGA GGAAGTGTGT
2201 TCTTTTACC CTGAAGTTAG TTGCATATTC AGAGGTAAAG TTGTGTGCTA
2251 TCTTGGCAGC ATCTTAGAGA TGGAGACATT AACAAAGCTAA TGGTAATTAG
2301 AATCATTTGA ATTTATTTT TTTCTAATATG TGAACACAGC ATTTCAAGTG
2351 TTTTATCTTT TTTTTTTTAA AATTTAAATG GGAATATAAC ACAGTTTTC
2401 CTTCATATT CCTCTCTTGA GTTTATGCAC ATCTCTATAA ATCATTAGTT
2451 TTCTATTTTA TTACATAAAA TTCTTTTAGA AAATGCAAAAT AGTGAACCTT
2501 GTGAATGGAT TTTTCCATAC TCATCTACAA TTCCTCCATT TTAATAGACT
2551 ACTTTTATTT TTTAATTTAA AAAATCTACT TCAGTATCAT GAGTAGGTCT
2601 TACATCAGTG ATGGGTTCTT TTTGTAGTGA GACATACAAA TCTGATGTTA

```

2651 ATGTTTGCTC TTAGAAGTCA TACTCCATGG TCTTCAAAGA CCAAAAAATG
2701 AGGTTTGTCT TTTGTAATCA GGAAAAAATA AATTAATGAA CCTTAAAAA
2751 AAAAAAATAA GG

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 105 bp to 1736 bp; peptide length: 544
Category: known protein
Classification: unclassified

1 MKHYSPTDYV NWLEEKVRQ KAGLEARKIV ASFSKRFFSE HVPCNGFSDI
51 ENLEGPEIFF EDELVCILNM EGRKALTWKY YAKKILYYLR QOKILNNLKA
101 FLQPPDDYES YLEGAVYIDQ YCNPLSDISL KDIQAQIDSI VELVCKTLRG
151 INSRHPSLAF KAGESSMIME IELQSQVLD MNYVLYDQLK FKGNRMDYYN
201 ALNLYMHQVL IRRGTGIPISM SLLYLTIAHQ LGVPLEPVNF PSHFLLRWQ
251 GAEGATLDIF DYIYIDAFGK GKQLTVKECE YLIGQHVTAA LYGVVNVKKV
301 LQRMVGNLLS LGKREGIDQS YQLLRDSL DL YLAMYPDQVQ LLLQARLYF
351 HLGWPEKSF CLVLKVLDDL QHIQTLDPGQ HGAVGYLVQH TLEHIERKKE
401 EVGVEVKLRS DEKHRDVCYS IGLIMKHKRY GYNCVIYGDW PTCMMGHEWI
451 RNMNVHSLPH GHHPFFYNVL VEDGSCRYAA QENLEYNVEP QEISHPDVGR
501 YFSEFTGTHY IPNAELEIRY PEDLEFVYET VQNIYSAKKE NIDE

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_8g5, frame 3

TREMBLNEW:AB020682_1 gene: "KIAA0875"; product: "KIAA0875 protein";
Homo sapiens mRNA for KIAA0875 protein, partial cds., N = 1, Score =
2832, P = 5.5e-295

>TREMBLNEW:AB020682_1 gene: "KIAA0875"; product: "KIAA0875 protein"; Homo
sapiens mRNA for KIAA0875 protein, partial cds.
Length = 621

HSPs:

Score = 2832 (424.9 bits), Expect = 5.5e-295, P = 5.5e-295
Identities = 537/544 (98%), Positives = 537/544 (98%)

Query: 1 MKHYSPTDYVNWLEEKVRQKAGLEARKIVASFSKRFFSEHVPCNGFSDIENLEGPEIFF 60
MKHYSPTDYVNWLEEKVRQKAGLEARKIVASFSKRFFSEHVPCNGFSDIENLEGPEIFF
Sbjct: 85 MKHYSPTDYVNWLEEKVRQKAGLEARKIVASFSKRFFSEHVPCNGFSDIENLEGPEIFF 144

Query: 61 EDELVCILNMEGRKALTWKYYAKKILYYLRQOKILNNLKAFLQPPDDYESYLEGAVYIDQ 120
EDELVCILNMEGRKALTWKYYAKKILYYLRQOKILNNLKAFLQPPDDYESYLEGAVYIDQ
Sbjct: 145 EDELVCILNMEGRKALTWKYYAKKILYYLRQOKILNNLKAFLQPPDDYESYLEGAVYIDQ 204

Query: 121 YCNPLSDISLKDIQAQIDSIVELVCKTLRGINSRHPSLAFKAGESSMIMEIELQSQVLD 180
YCNPLSDISLKDIQAQIDSIVELVCKTLRGINSRHPSLAFKAGESSMIMEIELQSQVLD
Sbjct: 205 YCNPLSDISLKDIQAQIDSIVELVCKTLRGINSRHPSLAFKAGESSMIMEIELQSQVLD 264

Query: 181 MNYVLYDQLKFKGNRMDYYNALNLYMHQVLIRRTGIPISMSLLYLTIAHQ LGVPLEPVNF 240
MNYVLYDQLKFKGNRMDYYNALNLYMHQVLIRRTGIPISMSLLYLTIAHQ LGVPLEPVNF
Sbjct: 265 MNYVLYDQLKFKGNRMDYYNALNLYMHQVLIRRTGIPISMSLLYLTIAHQ LGVPLEPVNF 324

Query: 241 PSHFLLRWQCGAEGATLDIFDYIYIDAFGKGKQLTVKECEYLIGQHVTAAALYGVVNVKKV 300
PSHFLLRWQCGAEGATLDIFDYIYIDAFGKGKQLTVKECEYLIGQHVTAAALYGVVNVKKV
Sbjct: 325 PSHFLLRWQCGAEGATLDIFDYIYIDAFGKGKQLTVKECEYLIGQHVTAAALYGVVNVKKV 384

Query: 301 LQRMVGNLLSLGKREGIDQSYQLLRDSL DLYLAMYPDQVQLLLQARLYFHLGIWPEKSF 360
LQRMVGNLLSLGKREGIDQSYQLLRDSL DLYLAMYPDQVQLLLQARLYFHLGIWPEK

Sbjct:	385	LQRMVGNLLSLGKREGIDQSYQLLRSLDLYLAMYPDQVQQLLLQARLYFHLGIWPEK--	442
Query:	361	CLVLKVLDIQLHIQTLDPGQHGAVGYLVOHTLEHIERKKEEVGVEVKLRSEKHHRDVCYS	420
		VLDILQHIQTLDPGQHGAVGYLVOHTLEHIERKKEEVGVEVKLRSEKHHRDVCYS	
Sbjct:	443	-----VLDILQHIQTLDPGQHGAVGYLVOHTLEHIERKKEEVGVEVKLRSEKHHRDVCYS	497
Query:	421	IGLIMKHKRYGYNVCVIYGDWPTCMMGHEWIRNMNVHSLPHGHHPFFYNVLVEDGSCRYAA	480
		IGLIMKHKRYGYNVCVIYGDWPTCMMGHEWIRNMNVHSLPHGHHPFFYNVLVEDGSCRYAA	
Sbjct:	498	IGLIMKHKRYGYNVCVIYGDWPTCMMGHEWIRNMNVHSLPHGHHPFFYNVLVEDGSCRYAA	557
Query:	481	QENLEYNVEPQEISHPDVGRYFSEFTGTHYIPNAELEIRYPEDLEFVYETVQNIYSAKKE	540
		QENLEYNVEPQEISHPDVGRYFSEFTGTHYIPNAELEIRYPEDLEFVYETVQNIYSAKKE	
Sbjct:	558	QENLEYNVEPQEISHPDVGRYFSEFTGTHYIPNAELEIRYPEDLEFVYETVQNIYSAKKE	617
Query:	541	NIDE	544
		NIDE	
Sbjct:	618	NIDE	621

Pedant information for DKFZphtes3 8q5, frame 3

Report for DKFZphtes3 8q5.3

```
[LENGTH]      544
[MW]           63307.22
[pI]           5.82
[HOMOL]        TREMBL:AB020682_1 gene: "KIAA0875"; product: "KIAA0875 protein"; Homo sapiens
mRNA for KIAA0875 protein, partial cds. 0.0
[KW]           Alpha Beta
[KW]           LOW COMPLEXITY      1.84 %
```

[illegible]

(No Prosite data available for DKFZphtes3 8q5.3)

(No Pfam data available for DKFZphtes3 8g5.3)

DKFZphtes3_8m10

group: nucleic acid management

DKFZphtes3_8m10 encodes a novel 221 amino acid protein with strong similarity to polyadenylate-binding proteins.

The poly(A)-binding protein (PABP) binds to the messenger (mRNA) 3'-poly(A) tail found on most eukaryotic mRNAs and together with the poly(A) tail has been implicated in governing the stability and the translation of mRNA.

The new protein can find application in modulation of mRNA translation and processing/stability.

strong similarity to polyadenylate-binding protein

frame shift at Bp 707-710

Sequenced by MediGenomix

Locus: unknown

Insert length: 2107 bp

Poly A stretch at pos. 2052, polyadenylation signal at pos. 2033

```
1 CGGAAAGGTC GCGGCTTGTC TGCCTGCGGG CAGCCGTGCC GAGAATGAAC
51 CCCAGCACCC CCAGCTACCC AACGGCCTCG CTCTACGTGG GGGACCTCCA
101 CCCGACGTG ACTGAGGCGA TGCTCTACGA GAAGTTCAGC CCGGCAGGGC
151 CCATCCTCTC CATCCGGATC TGCAGGGACT TGATCACCAG CGGCTCCTCC
201 AACTACGGCT ATGTGAACCT CCAGCATACG AAGGACGCGG AGCATGCTCT
251 GGACACCATG AATTTTGATG TTATAAAGGG CAAGCCAGTA CGCATCATGT
301 GGTCTCAGCG TGATCCATCA CTTGGAAGAA GTGGAGTGGG CAACATATTC
351 GTTAAAAATC TGGATAAGTC CATTAAATAT AAAGCACTGT ATGATACAGT
401 TTTGCTTTT GGTAAACATC TTTCTGTGTA CGTGGTTTGT GATGAAATG
451 GTTCCAGGGG TTATGGATTT GTACACTTTG AGACACACGA AGCAGCTGAA
501 AGAGCTATTA AAAAAATGAA CGGAATGCTC CTAATGGTTC GCAAAGTATT
551 TGTGGACAA TTTAAGTCTC GTAAAGAACG AGAAGCTGAA CTTGGAGCTA
601 GGGCAAAAGA GTTCCCAAT GTTTACATCA AGAATTTTGG AGAAGACATG
651 GATGATGAGC GCCTTAAGGA TCTCTTTGGC AAGTTCGGGC CCGCCTTAAG
701 TGTGAATTAA TGACCGATGA AAGTGGAATA TCCAAAGGAT TTGGATTTGT
751 AAGCTTTGAA AGGCATGAAG ATGCACAGAA AGCTGTAGAT GAGATGAATG
801 GAAAGGAGCT CAATGGAAAA CAAATTTACG TTGGTCGAGC TCAGAAAAAA
851 GTGGAACGGC AGACGGAACT TAAGCGCACA TTTGAACAGA TGAAGCAAGA
901 TAGGATCACC AGATACCAGG TTGTTAATCT TTATGTGAAA AATCTTGATG
951 ATGGTATTGA TGATGAACGT CTCGGGAAAG CGTTTTCTCC ATTTGGTACA
1001 ATCACTAGTG CAAAGGTTAT GATGGAAGGT GGTGCGCAGC AAGGGTTTGG
1051 TTTTGTATGT TTCTCCTCCC CAGAAGAAGC CACTAAAGCA GTTACAGAAA
1101 TGAACGGTAG AATTGTGGCC ACAAGCCAT TGTATGTAGC TTTAGCTCAG
1151 CGCAAAGAAG AGCGCCAGGC TTACCTCACT AACGAGTATA TGCAGAGAAT
1201 GGCAGTGTGA CGAGCTGTGC CCAACCAGCG AGCACCTCCT TCAGGTTACT
1251 TCATGACAGC TGTCCACAG ACTCAGAACC ATGCTGCATA CTATCCTCCT
1301 AGCCAAATTT CTCGACTAAG ACCAAGTCCT CGCTGGACTG CTCAGGGTGC
1351 CAGACCTCAT CCATTCCTAAA ATAAGCCAG TGCTATCCGC CCAGGTGCTC
1401 CTAGAGTACT ATTTAGTACT ATGAGACCAG CTTCTTCACA GGTTCACAGA
1451 GTCATGTCAA CGCAGCGTGT TGCTAACACA TCAACACAGA CAGTGGGTCC
1501 ACGTCCTGCA GCTGCTGCTG CTGCTGCAGC TACCCCTGCT GTGCGCACGG
1551 TTCCACGGTA TAAATATGCT GCGGGAGTTC GCAATCCTCA GCAACATCGT
1601 AATGCACAGC CACAAGTTAC AATGCAACAG CTTGCTGTTC ATGTACAAGG
1651 TCAGGAAACT TTGACTGCCT CCAGGTGGGC ATCTGCCCTT CCTCAAAAGC
1701 AAAAGCAAAT GTTAGGTGAA CGGCTCTTTC CTCTTATTCA AGCCATGCAC
1751 CCTACTCTTG CTGGGAAAAT CACTGGCATG TTGTTGGAGA TTGATAATTC
1801 AGAACTTCTT TATATGCTCG AGTCTCCAGA GTCACCTCCG TCTAAGGTTG
1851 ATGAAGCTGT AGCTGTACTA CAAGCCACCC AAGCTAAAGA GGCTACCCAG
1901 AAAGCAGTTA ACAGTGCTAC CGGTGTTCCA ACTGTTTAAA ATTGATCAGA
1951 GACCACGAAA AGAAATTTGT GCTTACCAGA AGAAAAATAT CTAACATCG
2001 AGAAACTATG GGAAAAAATC TTGCAAAATC TAAAAATAAA AATGCAAAAT
2051 CTAAAAATAA AAAAAAATAA AAAAAAATAA AAAAAAATAA AAAAAAATAA
2101 AAAAAGG
```

BLAST Results

Entry HSPOLYAB from database EMBL:
Human mRNA for polyA binding protein
Score = 5420, P = 0.0e+00, identities = 1162/1243

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 707 bp to 1936 bp; peptide length: 410
 Category: strong similarity to known protein
 Classification: unset
 Prosite motifs: RNP_1 (10-18)
 RNP_1 (112-120)

```

1 LMTDESGKSK GFGFVSFERH EDAQKAVDEM NGKELNGKQI YVGRAQKKVE
51 RQTELKRTFE QMKQDRITRY QVNLVYVKNL DDGIDDERLR KAFSPFGTIT
101 SAKVMEGGR SKGFGFVCFS SPEEATKAVT EMNGRIVATK PLYVALAQRK
151 EERQAYLTNE YMQRMASVRA VPNQRAPPSG YFMTAVPQTQ NHAAYYPPSQ
201 IARLRPSRW TAQGARPHPF QNKPSAIRPG APRVPFSTMR PASSQVPRVM
251 STQRVANTST QTVGPRPAAA AAAAATPAVR TVPRYKYAAG VRNPQHRNA
301 QPQVTMQQLA VHVQGQETLT ASRLASAPPQ KQKQMLGERL FPLIQAMHPT
351 LAGKITGMLL EIDNSELLEYM LESPELSRSK VDEAVAVLQA HQAKEATQKA
401 VNSATGVPTV

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_8m10, frame 2

PIR:DNHUPA polyadenylate-binding protein - human, N = 1, Score = 1931,
 P = 1.7e-199

PIR:I48718 poly(A) binding protein - mouse, N = 1, Score = 1928, P =
 3.6e-199

>PIR:DNHUPA polyadenylate-binding protein - human
 Length = 633

HSPs:

Score = 1931 (289.7 bits), Expect = 1.7e-199, P = 1.7e-199
 Identities = 384/415 (92%), Positives = 394/415 (94%)

```

Query:      1 LMTDESGSKGFGFVSFERHEDAQKAVDEMNGKELNGKQIYVGRAQKKVERQTELKRTFE 60
             +MTDESGSKGFGFVSFERHEDAQKAVDEMNGKELNGKQIYVGRAQKKVERQTELKR FE
Sbjct:    219 VMTDESGSKGFGFVSFERHEDAQKAVDEMNGKELNGKQIYVGRAQKKVERQTELKRKFE 278

Query:      61 QMKQDRITRYQVNLVYVKNLDDGIDDERLRKAFSPFGTITSKVMEGGRSKGFGFVCFS 120
             QMKQDRITRYQ VNLVYVKNLDDGIDDERLRK FSPFGTITSKVMEGGRSKGFGFVCFS
Sbjct:    279 QMKQDRITRYQVNLVYVKNLDDGIDDERLRKEFSFPGTITSKVMEGGRSKGFGFVCFS 338

Query:     121 SPEEATKAVTEMNGRIVATKPLYVALAQRKEERQAYLTNEYMQRMASVRAPVN-----Q 174
             SPEEATKAVTEMNGRIVATKPLYVALAQRKEERQA+LTN+YMQRMASVRAPVN      Q
Sbjct:    339 SPEEATKAVTEMNGRIVATKPLYVALAQRKEERQAHLTNQYMQRMASVRAPVNPVINYQ 398

Query:     175 RAPPSGYFMTAVPQTQNHAAYYPPSQIARLRPSRWTAQGARPHPFQNKPSAIRPGAPRV 234
             APPSGYFM A+PQTQN AAYYPPSQ+A+LRPSRWTAQGARPHPFQN P AIRP APR
Sbjct:    399 PAPPSGYMAAIPQTQNRAAYYPPSQVQLRPSRWTAQGARPHPFQNMMPGAIRPAAPRP 458

Query:     235 PFSTMRPASSQVPRVMSTQRVANTSTQTVGPRPAAAAAATPAVRTVPRYKYAAGVRNP 294
             PFSTMRPASSQVPRVMSTQRVANTSTQT+GPRPAAAAAATPAVRTVP+YKYAAGVRNP
Sbjct:    459 PFSTMRPASSQVPRVMSTQRVANTSTQTMGPRPAAAAAATPAVRTVPYKYAAGVRNP 517

Query:     295 QQHRNAQPQVTMQQLAVHVQGQETLTASRLASAPPQKQKQMLGERLFPLIQAMHPTLAGK 354
             QQH NAQPQVTMQQ AVHVQGQE LTAS LASAPPQ+QKQMLGERLFPLIQAMHPTLAGK
Sbjct:    518 QQHLNAQPQVTMQQPAVHVQGQEPLTASMLASAPPQEQKQMLGERLFPLIQAMHPTLAGK 577

Query:     355 ITGMLLEIDNSELLEYMLESPELSRSKVDEAVAVLQAHAQAKEATQKAVNSATGVPTV 410
             ITGMLLEIDNSELL+MLESPELSRSKVDEAVAVLQAHAQAKEA QKAVNSATGVPTV
Sbjct:    578 ITGMLLEIDNSELHLMLESPELSRSKVDEAVAVLQAHAQAKEAAQKAVNSATGVPTV 633

```

Score = 315 (47.3 bits), Expect = 1.9e-27, P = 1.9e-27

Identities = 71/163 (43%), Positives = 102/163 (62%)

Query: 1 LMTDESGKSGFGFVSFERHEDAQKAVDEMNGKELNGKQIYVGRAQKKVERQTELKRTFE 60
 ++ DE+G SKG+GFV FE E A++A++MNG LN +++VGR + + ER+ EL +
 Sbjct: 130 VVCDENG-SKGYGFVHFETQEAARAIEKMNGMLLNDRKVFVGRFKSRKEREAE LGARAK 188

Query: 61 QMKQDRITRYQVVNLYVKNLDDGIDDERLRKAFSPFGTITSAKVM-EGGRSKGFGFVCF 119
 + N+Y+KN + +DDERL+ F P S KVM E G+SKGFGFV F
 Sbjct: 189 EF-----TNVYIKNFGEDMDDERLKD LFGP---ALSVKVMTDESGKSGFGFVSF 235

Query: 120 SSPEEATKAVTEMNGRIVATKPLYVALAQRKEERQAYLTNEYMQ 163
 E+A KAV EMNG+ + K +YV AQ+K ERQ L ++ Q
 Sbjct: 236 ERHEDAQKAVDEMNGKELNGKQIYVGRAQKKVERQTELKRRKFEQ 279

Score = 214 (32.1 bits), Expect = 1.9e-14, P = 1.9e-14
 Identities = 50/150 (33%), Positives = 87/150 (58%)

Query: 8 KSKGFGFVSFERHEDAQKAVDEMNGKELNGKQIYVGRAQKKVERQTELKRTFEQMKQDRI 67
 +S G+ +V+F++ DA++A+D MN + GK + + +Q R L+++
 Sbjct: 50 RSLGYAYVNFQQPADAERALDTMNFVIRKGPVRIMWSQ----RDP SLRKS----- 96

Query: 68 TRYQVVNLYVKNLDDGIDDERLRKAFSPFGTITSAKVMMEGGRSKGFGFVCFSSPEEATK 127
 V N+++KNLD ID++ L FS FG I S KV+ + SKG+GFV F + E A +
 Sbjct: 97 ---GVGNIFIKNLDKSIDNKALYDTFSAFGNILSCKVCDENGSKGYGFVHFETQEAAR 153

Query: 128 AVTEMNGRIVATKPLYVALAQRKEERQAYL 157
 A+ +MNG ++ + ++V + ++ER+A L
 Sbjct: 154 AIEKMNGMLLNDRKVFVGRFKSRKEREAE L 183

Score = 120 (18.0 bits), Expect = 4.8e-04, P = 4.8e-04
 Identities = 30/99 (30%), Positives = 54/99 (54%)

Query: 70 YQVVNLYVKNLDDGIDDERLRKAFSPFGTITSAKVM--MEGGRSKGFGFVCFSSPEEATK 127
 Y + +LYV +L + + L + FSP G I S +V M RS G+ +V F P +A +
 Sbjct: 8 YPMASLYVGD LHPDVTEAMLYEKFSPAGPILSIRVCRDMITRRSLGYAYVNFQQPADAER 67

Query: 128 AVTEMNGRIVATKPLYVALAQRKEE-RQAYLTNEYMQRM 165
 A+ MN ++ KP+ + +QR R++ + N +++ +
 Sbjct: 68 ALDTMNFVIRKGPVRIMWSQRPDLRKS GVG NIFIKNL 106

Peptide information for frame 3

ORF from 45 bp to 707 bp; peptide length: 221
 Category: strong similarity to known protein
 Classification: unset
 Prosite motifs: RNP_1 (138-146)

1 MNPSTPSYPT ASLYVGD LHP DVTEAMLYEK FSPAGPILSI RICRDLITSG
 51 SSNYAYVNFQ HTKDAEHALD TMNFDVIK GK PVRIMWSQRD PSLRKS GVG N
 101 IFVKNLDKSI NNKALYDTVS AFGNILSCNV VCDENGSKGY GFVHFETHEA
 151 AERAIKKMNG MLLNGRKVFV GQFKSRKERE AELGARAKEF PNVIKNFGE
 201 DMDDERLKD L FGKFGPALSV N

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_8m10, frame 3

SWISSPROT:PAB1_HUMAN POLYADENYLATE-BINDING PROTEIN 1 (POLY(A) BINDING
 PROTEIN 1) (PABP 1)., N = 1, Score = 1039, P = 5.7e-105

PIR:I48718 poly(A) binding protein - mouse, N = 1, Score = 1031, P =
 4e-104

PIR:DNHUPA polyadenylate-binding protein - human, N = 1, Score = 1009,
 P = 8.7e-102

>SWISSPROT:PAB1_HUMAN POLYADENYLATE-BINDING PROTEIN 1 (POLY(A) BINDING
 PROTEIN 1) (PABP 1).
 Length = 636

HSPs:

Score = 1039 (155.9 bits), Expect = 5.7e-105, P = 5.7e-105
Identities = 199/220 (90%), Positives = 205/220 (93%)

Query: 1 MNPSTPSYPTASLYVGDLPDVTEAMLYEKFSFAGPILSIRICRDLITSGSSNYAYVNFQ 60
MNPS PSYP ASLYVGDLPDVTEAMLYEKFSFAGPILSIR+CRD+IT S YAYVNFQ
Sbjct: 1 MNPSAPSYPMASLYVGDLPDVTEAMLYEKFSFAGPILSIRVCRDMITRRSLGYAYVNFQ 60

Query: 61 HTKDAEHALDTMNFVIGKGPVRIMWSQRDPSLRKSGVGNIFVKNLDSINNKAALYDTVS 120
DAE ALDTMNFVIGKGPVRIMWSQRDPSLRKSGVGNIF+KNLDSI+NNKAALYDT S
Sbjct: 61 QPADAERALDTMNFVIGKGPVRIMWSQRDPSLRKSGVGNIFIKNLDSIDNKAALYDTFS 120

Query: 121 AFGNILSCNVVCDENGSKGYGFVHFETHEAAERAIAIKMNGMLLNDRKVFVVGQFQSRKERE 180
AFGNILSC VVCDENGSKGYGFVHFET EAAERAI+KMNGMLLN RKVFGV+FKSRKERE
Sbjct: 121 AFGNILSCKVVCDENGSKGYGFVHFETQEAARAIEKMNGMLLNDRKVFVVGFRFQSRKERE 180

Query: 181 AELGARAKEFPNVYIKNFGEDMDDERLKDLPFGKFGPALS 220
AELGARAKEF NVYIKNFGEDMDDERLKDLPFGKFGPALS
Sbjct: 181 AELGARAKEFTNVYIKNFGEDMDDERLKDLPFGKFGPALS 220

Score = 275 (41.3 bits), Expect = 4.1e-23, P = 4.1e-23
Identities = 71/233 (30%), Positives = 120/233 (51%)

Query: 2 NPSTPSYPTASLYVGDLPDVTEAMLYEKFSFAGPILSIRICRDLITSGSSNYAYVNFQ 61
+PS ++++ +L + LY+ FS G ILS ++ D S + + Q
Sbjct: 90 DPSLRKSGVGNIFIKNLDSIDNKAALYDTFSAFGNILSCKVVCDENGSKGYGFVHFETQE 149

Query: 62 TKD-AEHALDTMNFVIGKGPVRIMW-SQRDPSL--RKSGVGNIFVKNLDSINNKAALYD 117
+ A ++ M + K R +R+ L R N+++KN + ++++ L D
Sbjct: 150 AAERAIKEMNGMLLNDRKVFVVGFRFQSRKEREALGARAKEFTNVYIKNFGEDMDDERLKD 209

Query: 118 TVSAFGNILSCNVVCDENG-SKGYGFVHFETHEAAERAIAIKMNGMLLNDRKVFVVGQFQSR 176
FG LS V+ DE+G SKG+GFV FE HE A++A+ +MNG LNG++++VG+ + +
Sbjct: 210 LFGKFGPALSVMKMTDESGKSGFGFVSFERHEDAQKAVDEMNGKELNGKQIYVGRAQKK 269

Query: 177 KEREALGARAKEFP-----NVYIKNFGEDMDDERLKDLPFGKFGPALS 219
ER+ EL + ++ N+Y+KN + +DDERL+ F FG S
Sbjct: 270 VERQTELKRKFEQMKQDRITRYQGVNLYVKNLDDGIDDERLRKEFSFPGTITS 322

Score = 227 (34.1 bits), Expect = 6.3e-18, P = 6.3e-18
Identities = 57/187 (30%), Positives = 101/187 (54%)

Query: 12 SLYVGDLPDVTEAMLYEKFSFAGPILSIRICRDLITSGSSNYAYVNFQHTKDAEHALDT 71
++Y+ + D+ + L + F GP LS+++ D + S + +V+F+ +DA+ A+D
Sbjct: 192 NVYIKNFGEDMDDERLKDLPFGKFGPALSVMKMTDE-SGKSGFGFVSFERHEDAQKAVDE 250

Query: 72 MNFVIGKGPVRIMWSQR-----DPSLRKSGVGNIFVKNLDSINNKA 114
MN + GK + + +Q+ D R GV N++VKNLD I++
Sbjct: 251 MNGKELNGKQIYVGRAQKKVERQTELKRKFEQMKQDRITRYQGV-NLYVKNLDDGIDDER 309

Query: 115 LYDTVSAFGNILSCNVVCDENGSKGYGFVHFETHEAAERAIAIKMNGMLLNDRKVFVVGQFK 174
L S FG I S V+ + SKG+GFV F + E A +A+ +MNG ++ + ++V +
Sbjct: 310 LRKEFSFPGTITS AKVMMEGGRSKGFGFVCFSSPEEATKAVTEMNGRIVATKPLYVALAQ 369

Query: 175 SRKEREAL 183
++ER+A L
Sbjct: 370 RKEERQAH 378

Score = 100 (15.0 bits), Expect = 2.3e-02, P = 2.3e-02
Identities = 26/99 (26%), Positives = 53/99 (53%)

Query: 8 YPTASLYVGDLPDVTEAMLYEKFSFAGPILSIRICRDLITSG-SSNYAYVNFQHTKDAE 66
Y +LYV +L + + L ++FSP G I S ++ ++ G S + +V F ++A
Sbjct: 291 YQGVNLYVKNLDDGIDDERLRKEFSFPGTITS AKV---MMEGGRSKGFGFVCFSSPEEAT 347

Query: 67 HALDTMNFVIGKGPVRIMWSQRDPSLRKSGVGNIFVKNL 106
A+ MN ++ KP+ + +QR R++ + N +++ +
Sbjct: 348 KAVTEMNGRIVATKPLYVALAQRKEE-RQAHLTNQYMORM 386

Pedant information for DKFzphes3_8m10, frame 2

Report for DKFzphes3_8m10.2

[LENGTH] 409
[MW] 45235.68
[pI] 10.08
[HOMOL] SWISSPROT: PAB1_HUMAN POLYADENYLATE-BINDING PROTEIN 1 (POLY(A) BINDING PROTEIN 1) (PABP 1). 0.0

[FUNCAT] 04.05.05 mrna processing (5'-end, 3'-end processing and mrna degradation) [S. cerevisiae, YER165w] 1e-54
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YER165w] 1e-54
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YER165w] 1e-54
 [FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae, YER165w] 1e-54
 [FUNCAT] 04.05.99 other mrna-transcription activities [S. cerevisiae, YNL016w] 1e-15
 [FUNCAT] 11.01 stress response [S. cerevisiae, YGR159c] 1e-12
 [FUNCAT] 04.01.04 rna processing [S. cerevisiae, YGR159c] 1e-12
 [FUNCAT] 04.99 other transcription activities [S. cerevisiae, YNL175c] 4e-09
 [FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YPR112c] 5e-08
 [FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YHR086w] 3e-07
 [FUNCAT] 03.13 meiosis [S. cerevisiae, YHR086w] 3e-07
 [FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YHR086w] 3e-07
 [FUNCAT] 04.07 rna transport [S. cerevisiae, YOL123w HRPI - CF Ib] 9e-07
 [FUNCAT] 30.13 organization of chromosome structure [S. cerevisiae, YCL011c] 3e-06
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YGR250c] 8e-06
 [FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YDR432w] 2e-05
 [FUNCAT] 08.01 nuclear transport [S. cerevisiae, YDR432w] 2e-05
 [FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YFR023w] 3e-05
 [FUNCAT] 03.01 cell growth [S. cerevisiae, YBR212w] 3e-04
 [BLOCKS] BL00030B Eukaryotic RNA-binding region RNP-1 proteins
 [SCOP] dlsx1_4.34.7.1.3 Sex-lethal protein [(Drosophila melanogaster) 1e-17
 [PIRKW] nucleus 0.0
 [PIRKW] duplication 0.0
 [PIRKW] RNA binding 0.0
 [PIRKW] nucleolus 2e-09
 [PIRKW] tandem repeat 2e-09
 [PIRKW] single-stranded DNA binding 3e-06
 [PIRKW] DNA binding 5e-13
 [PIRKW] phosphoprotein 6e-10
 [PIRKW] ribosome 3e-08
 [PIRKW] mitochondrion 3e-08
 [PIRKW] alternative splicing 9e-11
 [PIRKW] chloroplast 2e-19
 [PIRKW] transcription regulation 2e-07
 [PIRKW] protein biosynthesis 3e-08
 [SUPFAM] nucleolin 6e-10
 [SUPFAM] glycine-rich RNA-binding protein 2e-07
 [SUPFAM] unassigned ribonucleoprotein repeat-containing proteins 2e-19
 [SUPFAM] polyadenylate-binding protein 0.0
 [SUPFAM] ribonucleoprotein repeat homology 0.0
 [PROSITE] RNP_1_2
 [PFAM] RNA recognition motif. (aka RRM, RBD, or RNP domain)
 [KW] Irregular
 [KW] 3D
 [KW] LOW_COMPLEXITY 5.62 %

SEQ MTDESGKSGFGFVSFERHEDAQKAVDEMNGKELNGKQIYVGRAQKKVERQTELKRTFEQ
 SEG
 1sxl-

SEQ MKQDRITRYQVNLVKNLDDGIDDERLRKAFSPFGTITSKVMMEGGRSKGFGFVCFSS
 SEG
 1sxl-CEEEECCCTTTTHHHHHHHHTTTTCCCCCEEECTTCTTTEEEECTTT

SEQ PEEATKAVTEMNGRIVATKPLYVALAQRKEERQAYLTNEYMQRMASVRAVPNQRAPPSGY
 SEG
 1sxl- HHHHHHHHHHTTTCCCCCBCCBCC.....

SEQ FMTAVPQTQNHAAAYPPSQIARLRPSPRWTAQGARPFPQNKPSAIRPGAPRVFSTMRP
 SEG
 1sxl-

SEQ ASSQVPRVMSTQRVANTSTQTVGPRPAAAAAATPAVRTVPRYKYAAGVRNPQQHRNAQ
 SEG
 1sxl-XXXXXXXXXXXXXXXXXXXXX.....

SEQ PQVTMQQLAVHVQGETLTASRLASAPPQKQKQMLGERLFPLIQAMHPTLAGKITGMLLE
 SEG
 1sxl-

SEQ IDNSELLYMLESPESLRKVDAAVAVLQAHQAKEATQKAVNSATGVPTV
 SEG
 1sxl-

Prosites for DKFZphtes3_8m10.2

PS00030	9->17	RNP_1	PDOC00030
PS00030	111->119	RNP_1	PDOC00030

Pfam for DKFZphtes3_8m10.2

HMM_NAME	RNA recognition motif. (aka RRM, RBD, or RNP domain)		
HMM	*IYVGNLPWDtTEEDLrDlFsQFGpivsIrMMrDReTGRSRGFVFEFED		
	+YV+NL+ ++E LR +FS+FG I+S+++M+ E GRS+GF+V F +		
Query	74	LYVKNLDDGIDDERLRKAfSPFGTITSAKVMM--EGGRSKGFGEVCFSS	120
HMM	EEDAekAideMNGmeFmGRrIRV*		
	+E+A+KA+ EMNG+++ ++++V		
Query	121	PEEATKAVTEMNGRIVATKPLYV	143

Pedant information for DKFZphtes3_8m10, frame 3

Report for DKFZphtes3_8m10.3

```

[LENGTH]      235
[MW]           26308.08
[pI]           8.95
[HOMOL]        SWISSPROT:PAB1_HUMAN POLYADENYLATE-BINDING PROTEIN 1 (POLY(A) BINDING PROTEIN
1) (PABP 1). 1e-113
[FUNCAT]       04.05.05 mRNA processing (5'-end, 3'-end processing and mRNA degradation) [S.
cerevisiae, YER165w] 1e-64
[FUNCAT]       30.03 organization of cytoplasm [S. cerevisiae, YER165w] 1e-64
[FUNCAT]       05.04 translation (initiation, elongation and termination) [S. cerevisiae,
YER165w] 1e-64
[FUNCAT]       30.10 nuclear organization [S. cerevisiae, YER165w] 1e-64
[FUNCAT]       03.19 recombination and DNA repair [S. cerevisiae, YFR023w] 1e-24
[FUNCAT]       11.04 DNA repair (direct repair, base excision repair and nucleotide excision
repair) [S. cerevisiae, YFR023w] 1e-24
[FUNCAT]       04.05.99 other mRNA-transcription activities [S. cerevisiae, YNL016w]
2e-19
[FUNCAT]       04.05.03 mRNA processing (splicing) [S. cerevisiae, YOR319w] 2e-14
[FUNCAT]       04.01.04 rRNA processing [S. cerevisiae, YGR159c] 1e-11
[FUNCAT]       11.01 stress response [S. cerevisiae, YGR159c] 1e-11
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YGR250c] 1e-09
[FUNCAT]       04.07 rRNA transport [S. cerevisiae, YOL123w HRP1 - CF Ib] 1e-09
[FUNCAT]       30.13 organization of chromosome structure [S. cerevisiae, YCL011c] 8e-09
[FUNCAT]       98 classification not yet clear-cut [S. cerevisiae, YPR112c] 2e-08
[FUNCAT]       03.13 meiosis [S. cerevisiae, YHR086w] 2e-08
[FUNCAT]       04.99 other transcription activities [S. cerevisiae, YBR212w] 3e-08
[FUNCAT]       03.01 cell growth [S. cerevisiae, YBR212w] 3e-08
[FUNCAT]       06.04 protein targeting, sorting and translocation [S. cerevisiae, YDR432w]
3e-04
[FUNCAT]       08.01 nuclear transport [S. cerevisiae, YDR432w] 3e-04
[BLOCKS]       BL00030B Eukaryotic RNA-binding region RNP-1 proteins
[BLOCKS]       BL00900D Bacteriophage-type RNA polymerase family proteins signatur
[SCOP]          d1sxl_ 4.34.7.1.3 Sex-lethal protein [(Drosophila melanogaster) 9e-23
[SCOP]          d2ula_ 4.34.7.1.2 U1A protein (human (Homo sapiens) 6e-24
[SCOP]          dlup1_2 4.34.7.1.1 Nuclear ribonucleoprotein A1, RNP A1, UP 1e-13
[PIRKW]         nucleus 1e-110
[PIRKW]         duplication 1e-110
[PIRKW]         RNA binding 1e-110
[PIRKW]         nucleolus 4e-10
[PIRKW]         tandem repeat 4e-10
[PIRKW]         single-stranded DNA binding 1e-06
[PIRKW]         DNA binding 9e-12
[PIRKW]         phosphoprotein 4e-10
[PIRKW]         mitochondrion 6e-07
[PIRKW]         heterotrimer 4e-06
[PIRKW]         alternative splicing 1e-15
[PIRKW]         chloroplast 5e-11
[PIRKW]         transcription regulation 3e-09
[PIRKW]         GTP binding 2e-06
[SUPFAM]        helix-destabilizing protein 1e-07
[SUPFAM]        nucleolin 4e-10
[SUPFAM]        glycine-rich RNA-binding protein 2e-07
[SUPFAM]        yeast HRP1 protein 2e-08

```

[SUPFAM] unassigned ribonucleoprotein repeat-containing proteins 3e-25
 [SUPFAM] polyadenylate-binding protein 1e-112
 [SUPFAM] ribonucleoprotein repeat homology 1e-112
 [PROSITE] RNP_1 1
 [PFAM] RNA recognition motif. (aka RRM, RBD, or RNP domain)
 [KW] All_Beta
 [KW] 3D

SEQ ERSRLVCLRAAVPRMNPSTPSYPTASLYVGDLPDVTEAMLYEKFSFAGPILSIRICRDL
 lhal-EEEEETTTTCHHHHHHHHGGGCCEEEEEEETT
 SEQ ITSGSSNYAYVNFQHTKDAEHALDTMNFVDVIKGPVRIMWSQRDPSLRKSGVGNIFVKNL
 lhal- TTTCEEEEEEEECCHHHHHHHHTTEE-TT---EEEEEECTTTCCCCCEEEEC
 SEQ DKSINNKAALYDTVSAFGNLSNVVCDENGSKGYGFVHFETHAAERAIKKMNGMLNLR
 lhal- TTTTCHHHHHHHHGGGCCEEEEEEETTTTCEEEEEECCHHHHHHH.....
 SEQ KVFVGQFKSRKEREALGARAKEFPNVYIKNFGEEDMDERLKDLPFGKFGPALSVN
 lhal-

Prosite for DKFZphtes3_8m10.3

PS00030 152->160 RNP_1 PDOC00030

Pfam for DKFZphtes3_8m10.3

HMM_NAME RNA recognition motif. (aka RRM, RBD, or RNP domain)
 HMM *IYVGNLPWDtTEEDLrDlFsQFGpIvsIrMMrDRtGRSRGFaFVEFED
 +YVG+L +D+TE +L + FS+ GPI+SIR+ RD T S +A+V+F+
 Query 27 LYVGDLPDVTEAMLYEKFSFAGPILSIRICRDLITSGSSNYAYVNFQH 75
 HMM EEDAekAIdemNGMeFmGRrIRV*
 DAE A+D+MN ++ G+++R+
 Query 76 TKDAEHALDTMNFVDVIKGPVRI 98
 HMM *IYVGNLPWDtTEEDLrDlFsQFGpIvsIrMMrDRtGRSRGFaFVEFED
 I+V+NL+ +++ L D S FG I+S++++ D + S+G++FV FE+
 Query 115 IFVKNLDKSINNKAALYDTVSAFGNLSNVVCD--ENGSKGYGFVHFET 161
 HMM EEDAekAIdemNGMeFmGRrIRV*
 +E+AE+AI +MNGM+++GR++ V
 Query 162 HEAAERAIKKMNGMLNLRKVFV 184

DKFZphtes3_8p7

group: testes derived

DKFZphtes3_8p7 encodes a novel 412 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

2 EST hits (both from testis librarys)

Sequenced by MediGenomix

Locus: unknown

Insert length: 2899 bp

Poly A stretch at pos. 2870, polyadenylation signal at pos. 2852

```
1 CCGACCCGCC CTGGGGTGCT GCGTGCGCTG CCTGCTCCCG CCTGAGGAAA
51 ACACTGCCCA TGGCGCAAGG CCGGGAGCGC GACGAAGGCC CCCACTCCGC
101 CGGCGGCGCG TCCTTGTCCTG TGAGATGGGT GCAAGGATTC CCTAAGCAGA
151 ATGTTCAATTT GTCAACGACA ACACCATTTG CTACCCCTGT GGAATTATG
201 TAATATTATTA TAATATTGAA ACCAAGAAAA AGACTGTACT GCAGTGTAGT
251 AATGGAATTG TGGGCGTCAT GGCAACTAAC ATCCCCTGTG AAGTTGTGGC
301 TTTTCTGAC CGGAAGCTAA AACCTCTCAT CTACGTATAC AGCTTCCAG
351 GATTGACCAG AAGGACCAAA TTGAAAGGCA ACATTCTCCT GGACTACACT
401 TTACTTTTCA TCAGTTACTG TGGCACCTAC CTGGCTAGTT ACTCCTCTCT
451 CCCAGAAATTT GAACCTGGCC TTTGGAACTG GGAATCGAGT ATCATTTTGT
501 GTAAGAAATC ACAGCCTGGA ATGGATGTGA ACCAAATGTC TTTTAACCCC
551 ATGAACCTGGC GCCAGCTGTG CTTATCAAGT CCAAGTACAG TGAGCGGTG
601 GACCATTGAA AGAAGTAACC AGGAGCATTG TTTCAGAGCA AGGTCGGTGA
651 AATTACCTCT AGAAGATGGG TCATTTTTTA ATGAACGGA TGTCGTTTTT
701 CCCAGTGTGT TGCCGAAAGA TCTCATCTAT GGTCCCGTGC TGCCACTGTC
751 AGCCATTGCC GGGCTGGTAG GCAAGAGGCG ACAGACTTTC CGGCCGAAAG
801 ATGATCTATA TCCTTTGCTT CACCCGACTA TGCATTGCTG GACTCCAACA
851 AGTGACTTGT ACATTGGCTG TGAAGAGGGT CATCTTTTAA TGATTAATGG
901 AGACACCTTG CAAGTGACTG TACTTAATAA GATAGAAGAG GAATCGCCAT
951 TGGAAGACAG AAGAAATTTT ATCAGTCCAG TAACCTTGGT ATATCAGAAG
1001 GAGGCGGTGC TGGCTTCTGG AATTGATGGC TTTGTGTATT CTTTATTAT
1051 TAAAGATAGA AGTTACATGA TCGAGGATTT TCTTGAGATT GAAAGACCTG
1101 TAGAACATAT GACATTTTCT CCCAATTATA CAGTGTGTCT GATTCAACA
1151 GACAAGGGAT CTGTTTATAT CTACACTTTT GGTAAAGGAG CAACCTTAAA
1201 TAAAGTCTTA GATGCTTGTG ATGGGAAATT TCAGGCAATT GACTTTATCA
1251 CACTTGGAAC CCAATACTTC ATGACACTTA CATATTCAGG GGAATTTGT
1301 GTTTGGTGGC TGGAGGATTG TGCTTGTTGA AGCAAGATTT ATCTGAATAC
1351 CCTAGCAACG GTTCTGGCTT GCTGTCCATC CTCCTCTCTG GCAGCCGTGG
1401 GCACGGAGGA TGGCTCGGTC TACTTCATCA GCGTATATGA TAAGGAATCC
1451 CCTCAGGTCT TGCACAAGGC CTTTCTCTCG GAATCGTCCG TGCAGCACGT
1501 CGTGTAAGTC CTTTCTGCCT CCAGGAGCGG CTCCTGTGTA CACCCGTCTG
1551 TTGAAATTTT TAGTGAAGCC ATCCTTTCTT TTAATTTTAA GTTTTACGTG
1601 TTTCAATTTG TTTGAATGTT AATATATTCA CACAGTTCAA CACTCAAAAG
1651 GTACAGAGGG CTGTGTAGTA AAGTACCCCT CATACCCAGG TCTGTCTTGT
1701 CAGGCAGCCT GGTACCAATT TCTCATGTCT CTCCTGAGAT GTTTTATCCA
1751 TGAACAAGCA AAACATAATA AGCACTTCTT TTTACTTGTA TCAATGGCCA
1801 TCATGTGTGT ATAGTGTGCC AGGCACTTCT GCTGTATTAA CTCCATGAGG
1851 TAAACACTCT TGTTGTCTCT ATTTGACAGG TGAGGAAGAT AAGGCACAAG
1901 GATTTTAAAT AACTTGCTCA ATAGTACACA GATAGTGAAT GGCAAAATGT
1951 GGGATTTGAA CCCAGGTAGT TGGGCTGCAG AGTCACTGCC TTTGCTCTTA
2001 AAAGGAGAAA ACTATGTACA ATGCCTCATT TCTTTTTCCT CTTAATCGTA
2051 TATCTTGGAG AATGTTTTAT ATCCACACAT AAAGACCAGC CTGATTATTT
2101 GTATAGCCAC ATAGTATTCC ATTATATGAA TATACTATCA TTTTAAATA
2151 ACGGTATATT AATGAACATT TAGAGTATTT CAAAACTTTT GAAGCAATAC
2201 TTTTAAGATG ATAATATAGA GACATTAGAT TTGGACTTGT AGGTGCTATC
2251 ATTATTACTG TTTCTTTTTA ATTTATTATA TTATTAGGTA TTAATAAGAA
2301 CAGACATTGT TATCTGCTT TACAGCTTGA GATCACTGTA GCTTGTGGCA
2351 TGTGATCTCT AAAACACCAG TCAGAAAGGT GTTATTCTTA TCCCTATTAG
2401 ACAAAATTAGG GAATTCAGGG TTAGAGAGGT GAGGAAAAGC ATTGTCCAAG
2451 ATTACACATT ACACAGCTAG CACACTGAGG AGCTGGCCCT GCCACTGTGG
2501 ACTGCCACAG TCCACCACCC TAGCTCAGTG GGAAGGATG GATAACCTCC
2551 TTCCATTTAC CCCCTGCCTT TCTGCACTGT CATTTTTTGT TGCCCTTCCCT
2601 TTCTCAGATC CTCTTATTCT AATTACATC TTCCCACTTT TTCTAATTGT
2651 ATAAAGTTGT AGACATGTTT CACTACATTC TTCTCCAC TGCCAGGTAC
2701 CAGACACAGG GTAATGAAAT GTCACACCCA CCACTAATTT GAGAATTGCT
```

2751 TATTTGCGCT TGAAACATCA AGAAAGCTCT ACCGACAGAC ATGTTTCATT
2801 CACTTATGAT GAACCAACTG CCCATCTTTA CTGAATCTTC TTGACTGTAT
2851 TTATTAAAGT TGCAATTTGG AAATAAAAAA AAAAAAAAAA AAAAAAAGG

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 269 bp to 1504 bp; peptide length: 412
Category: putative protein
Classification: no clue

1	MATNPICEVV	AFSDRLKLP	IYVYSFPGLT	RRTKLKGNIL	LDYTLLSFSY
51	CGTYLASYSS	LEPEFELANW	WESSILCKK	SQPGNQNVQM	SFNPMNSRQL
101	CLSSPSTVS	WTIERSNQEH	CFRARSVKLP	LEDGGSFNFT	DVVFPQSLPK
151	DLIYGPVLPL	SAIAGLVGKE	AETFRPKDDL	YPLLHPTMHC	WTPTSDLYIG
201	CEEGHLLMNI	GDTLQYTVLN	KTEEPSEST	RNRFTSPVTL	LYQKEGVLAS
251	GIDGFVYSFI	IKDRSMTVLE	FLEIERPVEH	MTFSPNYTVL	VIQTDKGSVY
301	IYTFGKEPTL	NKVLDACDCK	FQAFDFTPG	TQYFMTLTYS	GEICVWWLED
351	CACVSKSIYL	TLATVLACCP	SSLSAAVGTE	DGSVYFISVY	DKESPQVVKH
401	AFLSSESYOH	VV			

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3 8p7, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphtes3 8p7, frame 2

Report for DKFZphtes3_8p7.2

```
[LENGTH]      412
[MW]           46476.62
[pI]           4.91
[KW]           Alpha Beta
```

[illegible]

(No Prosite data available for DKFZphtes3_8p7.2)
(No Pfam data available for DKFZphtes3_8p7.2)

DKFZphtes3_9e22

group: testes derived

DKFZphtes3_9e22 encodes a novel 227 amino acid protein with weak partial similarity to Ring-finger proteins.

For the novel protein, Pfam, but not Prosite predicts a C3HC4 type RING finger motif. No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to zinc finger proteins

Sequenced by DKFZ

Locus: unknown

Insert length: 1318 bp

Poly A stretch at pos. 1308, no polyadenylation signal found

```
1 GCTCCCCCGG CTTTCGGAGC CCGGGGGCGG CCTGTGGCGC GCGGAGCCCG
51 CGCCGGAGCT CGCCTCTTTG GACCTTGAGG GGAACATGC GTTTCCTTTG
101 GATCGTTTGA AATTCTAAGT TTGGGATCCC CGCCCGCCCG CCTGCCTCTT
151 CCGCCCCGCG GGTTTTTTCC TTTTTCCTT TTGCTTTTTT TCCTTTTCTC
201 CCTCCGGGTC TCCTTTTTGA CTCCCTCCCC CTTTATGCTC GCCCAGCCCT
251 CCCCTGTGTC CTGAGAAAGT GGGGAGGGTC TCGGCCTCCA GGTTCGCCGC
301 CCACCGGGGC CCGGGCGAGC ATGGGGGGCA AGCAGAGCAC GCGGGCCCGC
351 TCCCGGGGCG CCTTCCCGGG GGTCTCCACC GATGACAGCG CCGTGCCGCG
401 GCCGGGAGGG GCGCCCCATT TCGGGCACTA CCGGACGGGC GCGGGGGCCA
451 TGGGGCTGCG CAGCCGCTCG GTCAGCTCGG TGGCAGGCAT GGGCATGGAC
501 CCCAGCACGG CCGGGGGGGT GCCCTTTGGC CTCTACACCC CCGCCTCCCG
551 GGGCACCAGC GACTCCGAGA GGGCGCCCGG CGGCGGAGGG TCTGCGTCCG
601 ACTCCACCTA TGCCCATGGC AATGGTTACC AGGAGACGGG CGGCGGTCAC
651 CATAGAGACG GGATGCTGTA CCTGGGCTCC CGAGCCTCGC TGGCGGATGC
701 TCTACCTCTG CACATCGCAC CCAGGTGGTT CAGCTCGCAT AGTGGTTTCA
751 AGTGCCCCAT TTGCTCCAAG TCTGTGGCTT CTGACGAGAT GGAATGCAC
801 TTTATAATGT GTTTGAGCAA ACCTCGCCTC TCCTACAACG ATGATGTGCT
851 GACTAAAGAC GCGGGTGAGT GTGTGATCTG CCTGGAGGAG CTGCTGCAGG
901 GGGACACGAT AGCCAGGCTG CCCTGCCTGT GCATCTATCA CAAAAGCTGC
951 ATAGACTCGT GGTTTGAAGT GAACAGATCT TGTCCGGAAC ACCCTGCGGA
1001 CTGACCTGCG GGCTTGCTTG CTGACTCCTC TCAAAGGGAC AGAGCGCCCC
1051 TGCTCCAGGG AGGAGGCTCA CCGGACCCTG GGGCAGAGCT GAGCTTGGGA
1101 CACCAGCGGG AACAGGGCAC CCCTTCTGCA CTGACTTCCA GATCATGGTT
1151 CTCCTTCTCT CCCTGAGGAC ACCAAATTGG ATGAGAGCAA GTTTGAGAGA
1201 AGAATGAATC AACTGCTATC CTTCCCTCA CCCCTCAGCC CAGGAGGGAA
1251 AGGGCATTTT CTTTTTCATC TTTGAAAGGC ATTGTGGGTC TGTCTTTAAA
1301 GTGTTTACAA AAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 321 bp to 1001 bp; peptide length: 227
Category: similarity to known protein
Classification: unclassified

```
1 MGGKQSTAAR SRGPFPGVST DDSAVPPPGG APHFGHYRTG GGAMGLRSRS
51 VSSVAGMGMD PSTAGGVVPG LYTPASRG TG DSERAPGGG SASDSTYAHG
101 NGYQETGGGH HRDGLMLYLG RASLADALPL HIAPRWFSH SGFKCPICSK
151 SVASDEMEMH FIMCLSKPRL SYNDVLT KD AGECVICLEE LLQGDITIAL
```

201 PCLCIYHKSC IDSWFEVNRSCPEHPAD

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_9e22, frame 3

TREMBL:AF078823.1 product: "RING-H2 finger protein RHA2b"; Arabidopsis thaliana RING-H2 finger protein RHA2b mRNA, complete cds., N = 1, Score = 111, P = 2.8e-06

TREMBL:AF078822.1 product: "RING-H2 finger protein RHA2a"; Arabidopsis thaliana RING-H2 finger protein RHA2a mRNA, complete cds., N = 1, Score = 112, P = 6.6e-06

TREMBL:AC004138.14 gene: "T17M13.17"; Arabidopsis thaliana chromosome II BAC T17M13 genomic sequence, complete sequence., N = 2, Score = 123, P = 1.4e-05

PIR:T02286 hypothetical protein T13D8.23 - Arabidopsis thaliana, N = 1, Score = 142, P = 8.8e-08

>PIR:T02286 hypothetical protein T13D8.23 - Arabidopsis thaliana
Length = 327

HSPs:

Score = 142 (21.3 bits), Expect = 8.8e-08, P = 8.8e-08
Identities = 24/57 (42%), Positives = 30/57 (52%)

Query: 166 SKPRLSYNDVLTQDAGECVICLEELLQGDITARLPCLCIYHKSCIDSWFEVNRSCP 222
S P + LT D +C +C+EE + G LPC IYHK CI W +N SCP
Sbjct: 206 SLPSVKITPQHLTNDMSQCTVCMEEFIVGGDATELPCKHIYHKDCIVPWLRLNNSCP 262

Pedant information for DKFZphtes3_9e22, frame 3

Report for DKFZphtes3_9e22.3

[LENGTH] 227
[MW] 23782.62
[pI] 6.18
[HOMOL] PIR:T02286 hypothetical protein T13D8.23 - Arabidopsis thaliana 2e-08
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YDR313c] 4e-06
[FUNCAT] 30.07 organization of endoplasmatic reticulum [S. cerevisiae, YOL013c]
0.001
[FUNCAT] 06.13 proteolysis [S. cerevisiae, YOL013c] 0.001
[PFAM] Zinc finger, C3HC4 type (RING finger)
[KW] Irregular

SEQ MGGKQSTAARSRGPFPGVSTDDSAVPPPGGAPHFHGYRTGGGAMGLRSRSVSSVAGMGMD
PRD ccc

SEQ PSTAGGVPFGLYTPASRGTDSEAPGGGSASDSTYAHNGYQETGGGHRDGMLYLGS
PRD ccc

SEQ RASLADALPLHIAPRWFSSHSFGFKPICSKSVASDEMEMHFIMCLSKPRLSYNDVLTQD
PRD hhhhhhhhhcecc

SEQ AGEVCVICLEELLQGDITARLPCLCIYHKSCIDSWFEVNRSCPEHPAD
PRD ccc

(No Prosite data available for DKFZphtes3_9e22.3)

Pfam for DKFZphtes3_9e22.3

HMM_NAME Zinc finger, C3HC4 type (RING finger)
HMM *CPICFctFQldyPWPfdePmMlPCgHsFCypCIrrW.....CPmC*
C IC L+++ D++ LPC+ ++ ++CI +W CP+
Query 184 CVIC-----LEELLQGDITARLPCLCIYHKSCIDSWFEVNRSCPEH 224

DKF2phtes3_9i20

group: testes derived

DKF2phtes3_9i20 encodes a novel 205 amino acid protein with similarity to human KIAA0336 gene.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: /map="44.1 cR from top of Chr17 linkage group"

Insert length: 2509 bp

Poly A stretch at pos. 2499, polyadenylation signal at pos. 2481

```
1 CTCGCCGAGA TGACCTGGGC ACCTCTGCGT TGAATCGGCA AATACTGATC
51 AAGCCGCATT TATTCTGCTC TCAGGAACCTC TAAGTCTAGC AGAGAAGATG
101 AGGCGGTAGA AGTTTCATCAA TGGCTTGGCT GGAGGACAAG CAAATTGAGG
151 ACATTGGCAA CGGAGTGATC AAAATGATAG ATCATGAGGC CTAAATGAA
201 TAAGGAAAGA AGAGAAGTGG CAGAGGCTGA GAACAGAAAC AGAGGGTGGA
251 GGGGCTGTAA ATCTTGAAGA TTAGGGTATA ATATGAGTAT ATGGGTAAAGA
301 ATTGGAAGAA TTGTGTAGGA GGCAGTAGTC AAAAAGTAGA AGCAGTTTGG
351 AAGAGTAGTT ACAAATATCA AGAGCCAGGT GGCTAAAAGG TGGAGCTATA
401 GGTCATTGAA GCTCAAGAAA CTGAGTCTCT AGGGCATTGG TTAAGTCATC
451 TGTCTAGACT TCAAAGTTGT CTAGGATGAT AATTCAGAAG ACTGATCTGT
501 GCCAAGTCA CAGGTTTTTC ACGACTGAAA ACAACATAGC AAAATAAGCC
551 AAGATGCTG TGGATCCAAT GACCTACGAG GCCCAGTTCT TTGGCTTCAC
601 GCCACAAACG TGCATGCTTC GGATCTACAT TGCATTTCAA GACTACCTAT
651 TTGAAGTGAT GCAGGCCGTT GAACAGGTTA TTCTGAAGAA GCTGGATGGC
701 ATCCAGACT GTGACATTAG CCCAGTGCAG ATTCGCAAT GCACAGAGAA
751 GTTCTTTGCT TTCAAGAAAG GACATTTTGA TAACCTTTTT AGCAAAATGG
801 AGCAACTGTT TTTGCAGCTG ATTTTACGTA TTCCCTCAA CATCTTGCTT
851 CCTGAAGATA AATGTAAGGA GACACCTTAT AGTGAGGAAG ATTTTCAGCA
901 TCTCCAGAAA GAAATGAAC AGTTACAGGA GAAGTACAAG ACTGAATTAT
951 GTACTAAGCA GGCCCTTCTT GCAGAATTAG AAGAGCAAAA AATTGTTTCA
1001 GCCAAACTCA AACAGACGTT GACTTTCTTT GATGAGCTTC ATAATGTTGG
1051 CAGAGATCAT GGGACTAGTG ATTTTAGGGA GAGTTTAGTA TCCCTGGTTC
1101 AGAAGTCCAG AAAACTACAG AACATTAGAG ACAATGTGGA AAAGGAATCG
1151 AAACCACTGA AAATATCTTA ATTGCTCAGT AGTCAAAAGG AGGAGCCTGT
1201 CAAAAGTAG AATCATAAGG ACTGTTCAAA CCATAAGGAC TGTTCAAATC
1251 ATACCACTGA CTGTTCAAA CACCATACT TTTTATTAGA TTTGCTTTGT
1301 CAACCTCTTC TTGTATTCTG TGTTTTCCTC TTTTGGGTC CACTTTGCTG
1351 AGGTATGAAG TGTACTACTT TGAAGTAGGC TGAAGCATCT GAGTCTTCTA
1401 ATAAGTGGA AGGGATCCAA CAAAGAAGCC ATGACCAGTT AAAGATATT
1451 GCAGAGTTAC ACCTTGGTCA TAAGTCCTTT GTGACCTTGA TTATTTTGGC
1501 TTAAGTCTTG GATGAGACCA GACAAGAAAA GGATTAAACG GGTGGCTCCT
1551 TTAATATTAT TATTATTGTT TTTGAGACAA GGTCCCTTTC TGTACCCAG
1601 GTTAGAGTAG ATTTCACTGG CACAATCTTG GCTCACTGCA ACCTCTGTGT
1651 CCTGGGCTCA AGTGATCCCT CTGCTCAGC CTCCCAAGTA GCTAGGACCA
1701 CAGGTGCGTG TCACCATGCT TGGCTAATTT TTTTGCAGAA ACGAGGCCCTC
1751 ACTATATTGT CCAGGCTGAG TGGCTCTTTT ATTAACCAAGT CATTACACTG
1801 CGGAACAGCC AACATAGAGT ACTTGCTCTC GTCCGTGTGA TTTTCTTTCA
1851 TGAGGGAGTC AATATGTAGT GGAAAGAAGC ATGTAGCAAA AAAGACAACC
1901 TTGATCTTTA ATAAAAAGA AGTTGGTTTA TTCCAAAAT AAATCCCCTG
1951 ACAAAAAACC TGGTGATGTT AAGCAATTGA CTGCTTTAGA GTCCAGCAGA
2001 AGACCTTAGA CAAAAAAGC AGAACCCTAG GGAGTAGAAA AGGAAGCATG
2051 TAGCATATAC TCAGTAGTGA AATTTAATTT TACTGACTGT TAGGTATCTA
2101 TGCCAATTTG TTTTCATACT TCAGTTGGTT TTGGAATCTG CCTTATACCT
2151 AATATTTATT TATTCACACT CATAAGCATC AAATATTTAA TGCCCTCAGT
2201 GGGAAATTTG TGTTTAAACT CAATGGAATC TAATATTTCT TTATGTCGTT
2251 AGTCCTCTGA AAATGTTAGG TCACCCAAGG AAAGGGGAGA AATAGCAATG
2301 GTTGTTCCTA AGGTATTGCT TGCCCTCCAT GTCTTCCTAA AGAGCAGAAC
2351 TTGGAGTTTC TCCTTTATGT AGAGAAGAAG TAACTTAGGG TGTATTTGCA
2401 ATGAAATATT CATAGATATT GAAAGCTTGT GTTTACATGA AATATGTTTA
2451 TTATCAAGAA GTCCTTTTTT CAATCTCTGA CATTAAATAT ATGTGTTTTA
2501 AAAAAAAA
```

BLAST Results

Entry AC004148 from database EMBL:
Homo sapiens chromosome 17, clone HCIT524C5, complete sequence.
Score = 5245, P = 0.0e+00, identities = 1049/1049
3 exons

Entry HS556361 from database EMBL:
human STS TIGR-A003N29.
Score = 1005, P = 1.3e-39, identities = 201/201

Entry HSG043 from database EMBL:
human STS SHGC-36031.
Score = 955, P = 2.8e-37, identities = 205/215

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 554 bp to 1168 bp; peptide length: 205
Category: putative protein
Classification: no clue

```

1 MSVDPMTYEA QFFGFTPQTC MLRIYIAFQD YLFEVMOAVE QVILKKLDGI
51 PDCDISPVQI RKCTEKFLCF MKGHFDNLFS KMEQLFLQLI LRIPSNILLP
101 EDKCKETPYS EEDFQHLQKE IEQLQEKYKT ELCTKQALLA ELEEQKIVQA
151 KKKQTITFFD ELHNVGRDHG TSDFRESLVS LVQNSRKLQN IRDNVEKESK
201 RLKIS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_9i20, frame 2

TREMBLNEW:HSAB2334_1 gene: "KIAA0336"; Human mRNA for KIAA0336 gene,
complete cds., N = 1, Score = 107, P = 0.0081

>TREMBLNEW:HSAB2334_1 gene: "KIAA0336"; Human mRNA for KIAA0336 gene,
complete cds.
Length = 1,583

HSPs:

Score = 107 (16.1 bits), Expect = 8.2e-03, P = 8.1e-03
Identities = 42/140 (30%), Positives = 76/140 (54%)

```

Query:   65 EKFLCFMKGHFDNLFSKMEQLFLQLILRIPSNILLPEDKCKETPYSEED----FQHLQKE 120
          EK  CF+K H +NL  +EQ  +L R   ILL +D  ++P + D   + L+++
Sbjct:   796 EKEKCFIKEH-ENLKPLLEQK--ELRDRRAELILL-KDSLAKSPSVKNDPLSSVKELEEK 851

Query:   121 IEQLQE--KYKTELCTKQALLAELEEQKIVQAKLQKTLTFDELHNVGRDHGTSDFRESL 178
          IE L++ K K E   K  L+A ++ +K + +  K+T T  +EL ++  +   S+
Sbjct:   852 IENLEKECKEKEEKINKIKLVA-VKAKKELDSSRKETQTVKEELESRLSEK--DQLSASM 908

Query:   179 VSLVQNSRKLQNIRDNVEKESKRLKI 204
          L+Q +   +N+   EK+S++L +
Sbjct:   909 RDLIQGAESYKNLLLEYEKQSEQLDV 934

```

Pedant information for DKFZphtes3_9i20, frame 2

Report for DKFZphtes3_9i20.2

```

[LENGTH]      205
[MW]           24140.13
[pI]           5.51
[KW]           All Alpha
[KW]           COILED_COIL      18.05 %

```

(No Pfam data available for DKFZphtes3 9i20.2)

DKFZphtes3_9k22

group: testes derived

DKFZphtes3_9k22 encodes a novel 304 amino acid protein with partial similarity to X. leavis katanin p80.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to C-terminus of katanin p80

Sequenced by DKF2

Locus: unknown

Insert length: 2676 bp

Poly A stretch at pos. 2665, no polyadenylation signal found

```
1 CTCTCTAGGC TGCCGGGCGC TGGTCGTCAG CGCCGAGGCT GGGCTGAGGC
51 GCCGCGGTAC CATGAGGCGC CGGTACTTAA GAGATTATGG CATCAGAAAC
101 CCACAAATGTT AAAAAACGGA ACTTTTGTA TAAGATTGAG GATCATTTC
151 TTGATCTTCC TAGAAAAAAG ATCTCTAATT TCACTAATAA GAACATGAAG
201 GAGGTTAAGA AATCTCCAAA ACAGTTGGCT GCTTACATAA ATAGAACAGT
251 TGGACAAACT GTGAAAAGCC CAGATAAACT TCGTAAAGTG ATCTATCGCA
301 GAAAGAAAGT TCATCATCCC TTTCCAAATC CTGTGTACAG AAAAAACAG
351 TCCCCTGGAA GTGGGGGCTG TGACATGGCA AATAAAGAAA ATGAACTGGC
401 TTGTGCAGGC CACCTGCCTG AAAAATTACA CCATGATAGT CGAACATATT
451 TGGTTAACTC CAGTGATTCT GGTCTTTCAC AGACAGAAAG CCCATCATCA
501 AAATATAGTG GGTTTTTTTC TGAGGTTTCT CAGGACCATG AAACAATGGC
551 CCAAGTTTGT TTCAGCAGGA ATATGAGATT GAATGTAGCT TTAACTTTCT
601 GGAGAAAGAG AAGTATAAGT GAACCTGTAG CTTATTTGTT GAGGATAGAA
651 GATCTTGGCG TTGTGGTAGA TTGCCTTCCT GTGCTCACCA ATTGTTTACA
701 GGAAGAAAAA CAATATATCT CACTTGGCTG CTGTGTGAC TTGTGCTC
751 TAGTAAAGTC ACTACTTAAA AGCAAAATTG AAGAATATGT TATAGTTGTT
801 TTAACCTGGC TTCAAGCAGT CATTAAGAGG TGGTGGTCAG AACTATCATC
851 CAAACAGAA ATATATAAATG ATGGAATAT TCAAAATTTA AAACAACAAT
901 TAAGTGGATT ATGGGAACAG GAAAACCATC TTACTTTGGT TCCAGGATAT
951 ACTGGTAATA TAGCTAAGGA TGTAGATGCT TATTTATTAC AGTTACATTG
1001 AGAGATTTCA TCTACTAAAG AGCATTGGT TTTTCAAAAC ATCCCTGAAC
1051 TGTATAATTT ACAAAAAATA AAGTCTCGTC TGAGAACTGT GAACTGTGGA
1101 AGAAATCAAA ACTATTTTTT CTTTTAAAAA GCCACGTAAT GAAACCACTA
1151 ATGAATATCCC AGCAATCTGC TTCACATTGA AGTGGAAAAA TATCCAAAAAG
1201 GAGCAGCTTC AATTTTCATTG AGGTGAAAGT GCACATGAA GATTGTTCAC
1251 CTTTGTGCGA TTTGGGAGTT ATATGGTTAT TTGGTAACAT TAAGAACTAC
1301 TGGATTTTAA TGCAATCCTG CATAAAAAATA TAATTTATAC TATGTGAAAA
1351 AATAAGACAG GACTTACCAC TAGGAACCA CAGAACCAAT CATCATTAAC
1401 TTTTTTAAGA TTGTGTTTTA TTAATAAAAA AAAACACTTA AATGTGTGCA
1451 GCTATTTTCT TATGTTGAAA AGACTGAAAG TTTAAACAT GAAAAAATC
1501 AATATTAAAC ATTTTTTGTT CACACTGAGA TACTGTGTAT GTAAAAATGCC
1551 TTAATTATTA ATAAGCCAAT GTGTTATGAT ACCAATATCT GTTTTAAAAA
1601 ACTAAACCA ACCATGCCTC TGGCATGATA AAATCATGGA ATTAAATCAG
1651 GGGTTTACAT TCTTGTAGAG TGTCTTGAA ACACCTCTCTG CACCATTTT
1701 AAAACTTGAG AATAGTTTTA GTATCTCTGA TATTTTTCG CAGAATCATC
1751 ATGTACATGA TGAATGTGTT ATCCCTATCT AAGGAAAAAG GTGAATATGT
1801 TTTTGTATGA ATGTTTAACT GGAATGTCC ATGGACTTGG CTAATTTATA
1851 TTTACTTTTT ATGTACATA GATTTCTAAT ATTTTTCATT CCTGTATCAT
1901 TTAACCTTCC TTCATTGAG TAAATCACT AAATATTCT ATTTTTTTCG
1951 TTTTTTAAAT TCTGATTTTA TATGAATCT AATCTTTTT CACTACATAT
2001 GTTTTAAAGA GTTACATACA GTGATTTAGA ATGGTTTACA GTTAATGCTG
2051 ATCTTGTATT TTAAATTCCA ACACCTTGTG TCACTACCTC CTCTAATGGT
2101 TAGTATGATA TGCTAGCAGA CTGTATGAGG TCTTTTTTTA AAATACCACT
2151 TTTAGTGTCA GTGAACCAAA TTCTGGAATG TCTTAACAGC TCTAAATCTT
2201 ACTTGCTTGG AAAATGATTG GGGTTTAATA CCACTGCTGG TGGTTCACAC
2251 ATCATCCCAT CCTTAATATG CCTGACAGGC ATCTGAGCAA AGGTTTTTAG
2301 TAATTGAATT TCTCTGCAGT AGTCCTTCAA GCACCTGAAT GTAAACCTTT
2351 AGCATTATTT CGTTTAATGA CTACTGATAC GAATCTCAAG CAGATTCTT
2401 GCTCTTAAAA GTTATGTTTC ACTGAGTTCT GGTTTTGTGT AGCTATATTT
2451 TATATAGCTA GATATTCCTC ACAGTGAACA TGAATTGTAA TAATTGGTTA
2501 TTTCTTAAAG TCTTTAGATT ATAATAATTT CAGATTATTG CACGCTGTG
2551 ATTTGAGAGG TGAGTTATTT AAGAGGCCAG TTTTCAGGAC ATGGGAATTT
2601 GAATTGTAAA CCTGTTATCT CTGTGAAACT TTTAACATGA TAAATATATA
2651 CCTTCTTTTG TGCTTAAAAA AAAAAA
```

BLAST Results

Entry HS541354 from database EMBL:
human STS WI-11840.
Score = 1267, P = 7.1e-50, identities = 271/281

Medline entries

98227670:
Katanin, a microtubule-severing protein, is a novel AAA ATPase
that targets to the centrosome using a WD40-containing subunit.

Peptide information for frame 3

ORF from 87 bp to 998 bp; peptide length: 304
Category: similarity to known protein
Classification: unclassified

```

1 MASETHNVKK RNECNKIEDH FIDLPRKKIS NFTNKNMKEV KKSPKQLAAY
51 INRTVGQTVK SPDKLRKVIY RRKKVHHPPF NPCYRKKQSP GSGGCDMANK
101 ENELACAGHL PEKLHDSRT YLVNSSDSGS SQTESPSSKY SGFFSEVSQD
151 HETMAQVLF SRNRLNVALT FWRKRSISEL VAYLLRIEDL GVVVDCLPVL
201 TNCLQEEKQY ISLGCCVDLL PLVKSLLKSK FEEYVIVGLN WLQAVIKRWW
251 SELSSKTEII NDGNIQILKQ QLSGLWEQEN HLTLPVGYTG NIAKDVDAYL
301 LQLH

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_9k22, frame 3

TREMBL:AF056021_1 product: "p80 katanin"; *Xenopus laevis* p80 katanin
mRNA, partial cds., N = 1, Score = 146, P = 1.2e-07

TREMBL:AF052432_1 product: "katanin p80 subunit"; *Homo sapiens* katanin
p80 subunit mRNA, complete cds., N = 1, Score = 150, P = 1.2e-07

TREMBL:AF052433_1 product: "katanin p80 subunit"; *Strongylocentrotus*
purpuratus katanin p80 subunit mRNA, complete cds., N = 2, Score = 146,
P = 4.2e-07

>TREMBL:AF052432_1 product: "katanin p80 subunit"; *Homo sapiens* katanin p80
subunit mRNA, complete cds.
Length = 655

HSPs:

Score = 150 (22.5 bits), Expect = 1.2e-07, P = 1.2e-07
Identities = 35/105 (33%), Positives = 55/105 (52%)

```

Query:  145 SEVSQDHETMAQVLF SRNRLNVALT FWRKRSISEL VAYLLRIEDL GVVVDCLPVL TNCL 204
      S++ + H+TM VL SR+ L+   W   I   V   + I DL VVVD L   N +
Sbjct:  489 SQIRKGHDTMCVVLTSRHNLDTVRAVWTMGDIKTSVDSAVAINDL SVVVDLL----NIV 544

Query:  205 QEEKQYISLGCCVDLLPLVKSLLKSKFEEYVIVGLNWLQAVIKRW 249
      ++   L C +LP ++ LL+SK+E YV G   L+ +++R+
Sbjct:  545 NQKASLWKLDLCTTVLPQIEKLLQSKYESYVQTGCTSLKLILQRF 589

```

Pedant information for DKFZphtes3_9k22, frame 3

Report for DKFZphtes3_9k22.3

```

[LENGTH]      304
[MW]           34767.24
[pI]           9.18
[KW]           All_Alpha

```

```
[KW]          LOW_COMPLEXITY      3.95 %

SEQ  MASETHNVKKRNFCNKIEDHFDLPRKKISNFTNKNMKEVKKSPKQLAAYINRTVGQTVK
SEG  .....
PRD  cccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhcccccc

SEQ  SPDKLRKVIYRRKKVHHFFPNPCYRKKQSPGSGGCDMANKENELACAGHLPEKLHHSRT
SEG  .....
PRD  ccchhhhhhhhhhhccccccccccccccccccccccccchhhhhhhccccccccccce

SEQ  YLVNSSDSGSSQTESPSSKYSGFFSEVSQDHETMAQVLF SRNMRLNVALTFWRKRSISEL
SEG  .....
PRD  eeeeeccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhh

SEQ  VAYLLRIEDLGVVVDCLPVLTNCLQEERQYISLGCCVDLLPLVKSLLKSKFEYVIVGLN
SEG  .....xxxxxxxxxxxxx.....
PRD  hhhhhhhhhccceeecccchhhhhhhceeeccceeehhhhhhhhhhheeeeeehh

SEQ  WLQAVIKRWSELSSKTEIINDGNIQILKQQLSGLWEQENHLTLVPGYTGNIAKDVDAYL
SEG  .....
PRD  hhhhhhhhhhhccceeeccccccccccccccccchhhhhhhhhccccccccchhhhhhhh

SEQ  LQLH
SEG  ....
PRD  hccc
```

(No Prosite data available for DKFZphtes3_9k22.3)

(No Pfam data available for DKFZphtes3_9k22.3)

Localization of expressed proteins

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
DKFZp434B0435	AL117496	6248	48	5366	1773	transport and traffic	similar to: kinesin like proteins	512.1 cR from top of Chr10 linkage group	"secr pathway"	None
DKFZp434N0535	AL117518	4055	126	4025	1300	differentiation & development	similar to: Drosophila chromatin protein		"no predict"	None
DKFZp564A0122	AL110209	2722	65	1300	412	signaling & communication	similar to: acyltransferase	16	"mitochondria"	Mitochondria
DKFZp564A022	AL136620	1376	132	632	167	unknown	unknown	4	"no predict"	Endoplasmic Reticulum
DKFZp564A032	AL50267	2214	76	1953	626	differentiation & development	similar to: MG21 contains three conserved protein motifs present in GTP-binding proteins, but these are not conserved in 2_2a3.1	238.7 cR from top of Chr20 linkage group	"no predict"	Nucleus
DKFZp564A0723	AL80116	2524	42	2177	712	cell cycle	similar to: origin recognition complex	6q14.3-16.1	"nucleus"	Cytosol + Nucleus
DKFZp564A202	AL80056	707	33	620	196	metabolism	similar to: protein involved in energy metabolism	10	"mitochondria"	Endoplasmic Reticulum
DKFZp564B0482	AL110243	2092	317	1579	421	signaling & communication	Unknown, contains 2 WD-40 repeats, which are typical for the beta-transducin subunit of G-proteins		"no predict"	Cytosol + Nucleus
DKFZp564B1023	AL136611	2905	157	1896	580	nucleic acid management	similar to: RNA helicase		"nucleus / nuclear envelope"	Nucleus

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
DKFZp564B1162	AL136646	4593	661	2625	655	signaling & communication	unknown	4	"no predict"	Cytoskeleton
DKFZp564B122	AL049972	1842	70	1536	488	unknown	unknown		"no predict"	Cytosol
DKFZp564B1471	AL136667	1484	78	323	82	membrane protein	unknown		"no predict"	Endoplasmic Reticulum
DKFZp564B162	AL136621	1914	246	1631	462	nucleic acid management	similar to: Zinc finger protein	13q12	"no predict"	other/unknc
DKFZp564B163	AL50268	1208	191	577	129	transport and traffic	similar to: GTP binding protein		"secre pathway"	Cytosol + Nucleus
DKFZp564B212	AL136623	1915	218	1348	377	protein management	similar to: protein involved in posttranslational modification	22q12.1	"secre pathway"	Endoplasmic Reticulum
DKFZp564B2123	AL136612	3300	121	699	193	signaling & communication	similar to: Neurocalcin is a Ca(2+)-binding protein with 3 EF-hands. Homology with recoverin indicates involvement in Ca2+ dependent activation of guanylate cyclase.	574.6 cR from top of Chr8 linkage group	"no predict"	Cytosol + Nucleus
DKFZp564B246	AL136664	2054	73	1074	334	metabolism	similar to: dTDP-6-deoxy-L-mannose-dehydrogenase	5	"nucleus or cytosol"	Nucleus
DKFZp564C0362	AL80076	1731	60	1142	361	nucleic acid management	similar to: ssDNA binding protein		"no predict"	other/unknown
DKFZp564C0469	AL050298	899	86	898	270	unknown	unknown		"nucleus"	Cytosol + Nucleus
DKFZp564C1362	AL136647	837	137	673	179	metabolism	similar to: molecular clock protein	16p12.3-p13.11	"mitochondria"	Other/unknown

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
DKFZp564C1616	AL136597	3928	240	1997	586	structure & motility	shares the features of mayven and kelch and therefore should be involved in the organisation of cytoskeleton binding to membrane proteins		"cytoskeleton / plasma membrane"	Nucleus
DKFZp564C162	AL136627	2305	155	625	157	membrane protein	unknown	86.2 cR from top of Chr1 linkage group	"no predict"	Endoplasmic Reticulum
DKFZp564C1664	AL136656	1866	180	1040	287	unknown	unknown	745_A_2; 756_F_2; 842_C_2	"no predict"	Cytosol
DKFZp564C182	AL136628	2835	272	1177	302	unknown	unknown		"no predict"	Golgi
DKFZp564C183	AL136639	1709	105	1448	448	nucleic acid management	similar to: DEAD-box helicase	87.50 cR from top of Chr16 linkage group	"nucleus / nuclear envelope"	Nuclear envelope
DKFZp564C196	AL050020	2266	366	966	200	signaling & communication	similar to: neuronal calcium sensor		"no predict"	Nucleus
DKFZp564D116	AL050022	2535	29	1849	607	signaling & communication	similar to: GTP-binding protein		"no predict"	Cytosol
DKFZp564D202	AL136631	1787	18	944	309	unknown	unknown		"no predict"	Cytosol
DKFZp564E0123	AL136613	2005	104	1000	299	unknown	unknown	16q13	"no predict"	Cytosol + Nucleus
DKFZp564E0482	AL136697	2923	163	1581	473	signaling & communication	similar to: calmodulin-related protein	200.5 cR from top of Chr3 linkage group	"membranes"	Cytosol + Nucleus

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
DKF2p564E1782	AL136696	1618	40	972	311	membrane protein	unknown	171.7 cR from top of Chr14 linkage group	"no predict"	Endoplasmic Reticulum
DKF2p564E2182	AL50261	2367	193	804	204	Cell Cycle	similar to: protein involved in cell cycle, DNA repair, maintenance of minichromosomes	6q22.1-22.33	"nucleus"	Nucleus
DKF2p564F0223	AL136614	1016	68	613	182	unknown	unknown	12q24	"secretory pathway / endosomes"	other/unknown
DKF2p564F052	AL049989	1649	34	1303	423	signaling & communication	similar to: sorting nexin 7		"membranes"	Cytosol
DKF2p564F0522	AL049943	2078	283	943	220	unknown	unknown	2	"no predict"	Nucleus
DKF2p564F1862	AL80081	1987	250	918	223	differentiation & development	similar to: DnaJ proteins, but lacks CRR domain of these proteins.	7q31	"no predict"	Endoplasmic Reticulum
DKF2p564F2116	AL136598	1512	115	738	208	membrane protein	unknown	15q25	"nucleus"	other/unknown
DKF2p564F2122	AL136604	1910	156	1856	567	unknown	unknown	311.4 cR from top of Chr14 linkage group	"no predict"	Cytoskeleton (microtubules)
DKF2p564F2162	AL136648	1549	95	730	212	unknown	unknown	209.8 cR from top of Chr20 linkage group	"peroxisomes"	Peroxisomes
DKF2p564G0222	AL80115	1165	157	933	259	nucleic acid management	unknown		"no predict"	Endoplasmic Reticulum

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
DKF2p564G083	AL136641	1027	37	570	178	protein management	similar to: yeast, ARD1 and NAT1, are required for the expression of an N-terminal protein acetyltransferase 1.	20	"no predict"	Cytosol + Nucleus
DKF2p564G182	AL136632	2444	539	1225	229	unknown	unknown	6p22.1-22	"no predict"	Cytosol + Nucleus
DKF2p564H012	AL136633	957	93	632	180	unknown	unknown		"no predict"	Mitochondri
DKF2p564H1122	AL136605	1734	159	1133	325	membrane protein	unknown	11q14	"no predict"	Nucleus
DKF2p564H1322	AL136606	2292	270	1829	520	membrane protein	unknown	19q13.2 from BCKDHA-D19S217	"no predict"	Cytosol + Nucleus
DKF2p564H1562	AL136649	2014	75	971	299	structure & motility	similar to: Cell cell interaction protein	1	"plasma membrane"	Plasma membrane + cell contact sites
DKF2p564I0123	AL136615	1467	126	1064	313	signaling & communication	similar to: protein activator of the interferon-induced protein kinase		"cytosol or nucleus"	Cytosol
DKF2p564I0422	AL136607	4748	511	1194	228	signaling & communication	unknown		"no predict"	Golgi + Plasma membrane
DKF2p564I1216	AL136600	1548	81	635	185	membrane protein	unknown	873.3-875.1 CR from top of Chr1 linkage group	"no predict"	Endoplasmic Reticulum
DKF2p564I1782	AL136699	1741	168	410	81	signaling & communication	similar to: phospholemma protein,	11q23	"secre pathway"	Golgi + Plasma

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localisation Predicted	Localization
							a membrane substrate for the CAMP-dependent protein kinase; seems to serve as chloride channels or as chloride-channel regulators. Transmembrane Protein			membrane
DKF2p564J1206	AL136665	1122	34	921	296	unknown	unknown	377.5 cR from top of Chr8 linkage group	"mitochondria"	Mitochondria
DKF2p564J12423	AL136616	1713	58	882	275	metabolism	similar to: protein involved in amino acid metabolism	8p11.2	"cytosol"	Cytosol + Nucleus
DKF2p564J12482	AL136700	1860	10	1650	547	nucleic acid management	similar to: Dead-box helicase	175.5 cR from top of Chr7 linkage group	"nucleus / nuclear envelope"	Nucleus
DKF2p564J1022	AL110301	1409	5	1021	290	nucleic acid management	Unknown, contains a Leucine zipper	12	"cytosol or nucleus"	Cytosol + Nucleus
DKF2p564J1516	AL136601	2868	352	1839	496	structure & motility	similar to: RNA binding, Tubulin binding	20, 12.10 cR from GCT10F11	"cytosol"	Cytosol
DKF2p564J1864	AL136660	690	109	648	180	transport and traffic	similar to: canin and chicken microsomal signal peptidase 23 kd subunit.		"endoplasmic reticulum"	Endoplasmic Reticulum
DKF2p564J2222	AL136608	1858	154	1440	429	structure & motility	similar to: actin-related protein		"plasma membrane / cytoskeleton"	Plasma membrane
DKF2p564K0322	AL136609	2775	779	2392	538	unknown	unknown		"no"	Plasma

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
									predict"	membrane
DKFZp564K0822	AL136610	2789	10	525	172	unknown	unknown	7	"no predict"	Golgi
DKFZp564K1216	AL49933	1938	357	1418	354	signaling & communication	similar to: GRP-binding regulatory protein	7	"membranes"	Golgi + Plasma membrane
DKFZp564K192	AL136637	1931	107	1015	303	unknown	unknown	6p22.1-22.3	"no predict"	Other/unknown
DKFZp564K1964	AL117619	1560	207	884	226	unknown	unknown	17	"no predict"	Endoplasmic Reticulum
DKFZp564K2216	AL136602	2088	832	1155	108	unknown	unknown		"no predict"	Mitochondria
DKFZp564L023	AL136643	2978	279	2045	589	protein management	Unknown, Pfam prediction: ubiquitin family	9	"cytosol"	Cytosol + Nucleus
DKFZp564L1216	AL136603	2042	73	873	267	membrane protein	unknown		"secr pathway"	Golgi + plasma membrane
DKFZp564L2423	AL136617	2416	29	1072	348	transport and traffic	Unknown, a lectin character is predicted	2	"endoplasmic reticulum"	Endoplasmic Reticulum
DKFZp564M082	AL80071	902	227	589	121	unknown	Unknown, contains osteopontin motive		"no predict"	Cytosol + Nucleus
DKFZp564M112	AL80070	2686	14	595	194	signaling & communication	unknown	956.7 cR from top of Chr2 linkage group	"no predict"	Golgi
DKFZp564M173	AL136644	636	26	400	125	unknown	similar to: janus proteins		"no predict"	Cytosol + Nucleus
DKFZp564M1863	AL117602	1192	125	1027	301	signaling &	similar to: phospho-ducin-like protein, G-protein	9	"cytosol"	Cytosol

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation	Localization Predicted	Localization
						communication	modulator			
DKFZp564M1982	AL390217	2707	302	1160	286	unknown	unknown		"no predict"	Cytosol
DKFZp564M2423	AL80119	2201	86	1246	387	unknown	unknown	72.60 cR from top of Chr3 linkage group	"cytosol"	Cytosol
DKFZp564N0582	AL50264	1646	75	506	144	cell cycle	similar to: DRR1 gene	3p21.1	"cytoskeleton / plasma membrane"	Cytoskeleton (focal adhesion sites) + nucleus
DKFZp564N1623	AL136618	2936	172	1047	292	signaling & communication	Unknown, contains a WW domain which binds proteins with particular proline-motifs, [AP]-P-P-[AP]-Y, and thus resembles somewhat SH3 domains. This domain is frequently associated with other domains typical for proteins in signal transduction processes		"no predict"	Cytosol + Nucleus
DKFZp564O043	AL050390	2515	186	1509	441	structure & motility	similar to: ankyrin	7	"no predict"	Cytosol + Nucleus
DKFZp564O0523	AL136619	1736	24	1103	360	unknown	unknown	7q21-q22	"no predict"	Nucleus
DKFZp564O123	AL80122	1985	234	872	213	unknown	unknown		"no predict"	Cytosol + Nucleus
DKFZp564O1762	AL136652	1260	56	901	282	signaling & communication	similar to: low-density lipoprotein (LDL) receptors are the major		"secretory pathway"	Golgi

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
							cholesterol-carrying lipoproteins of plasma. The novel protein contains an additional leucine zipper suitable for protein-protein interaction.			
DKFZp56401923	AL050295	2091	237	2090	617	metabolism	similar to: dTDP-6-deoxy-L-mannose dehydrogenase		"secre pathway"	Cytosol
DKFZp56402423	AL390214	3564	656	1072	139	unknown	Unknown, contains CAAX box (prenyl group binding site); found in :Ras proteins, and ras-like proteins such as Rho, Rab, Rac, Ral, and Rap; nuclear lamins A and B; Some G protein alpha subunits, G protein gamma subunits; some dnaj-like proteins		"no predict"	Cytosol + Nucleus
DKFZp5640243	AL050015	1074	23	834	270	unknown	unknown	3	"no predict"	Endoplasmic Reticulum
DKFZp56611024	AL050037	1783	5	970	322	unknown	similar to: hypothetical protein Rv0712 - Mycobacterium tuberculosis		"no predict"	Cytosol
DKFZp566J2046	AL136720	1706	16	678	221	metabolism	similar to: 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase.	16	"no predict"	Mitochondria
DKFZp566K144	AL136727	3084	456	1079	208	transport and traffic	similar to: Rab protein		"secre pathway"	Golgi
DKFZp586D0919	AL050100	2777	48	494	148	unknown	unknown	12	"no"	Golgi

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
									"predict"	
DKFZp586E1124	AL136942	2005	184	861	226	transport and traffic	similar to: golgi transmembrane spanning transporter	8	"Golgi"	Golgi + plasma membrane
DKFZp586E1323	AL136936	1854	367	954	196	protein management	similar to: heat shock protein	578.9 cR from top of Chr12 linkage group	"cytosol or nucleus"	Cytosol + Nucleus
DKFZp586E1519	AL050101	2140	82	1680	559	unknown	similar to: A.thaliana A_IG002N01		"no predict"	Cytosol
DKFZp586F1918	AL050091	3489	184	594	137	unknown	unknown		"no predict"	Cytosol + Nucleus
DKFZp586F1919	AL136915	2024	134	745	204	membrane protein	unknown	14.8 cR from top of Chr20 linkage group	"secre pathway"	Golgi + plasma membrane
DKFZp586H2219	AL50282	1971	199	1623	475	unknown	unknown	22q11.2-qter	"no predict"	Cytosol
DKFZp586I0418	AL136912	1568	163	822	220	unknown	unknown	7q31	"no predict"	Cytosol + Nucleus
DKFZp586I1520	AL050149	2439	11	1711	566	transport and traffic	similar to: nuclear RanGTP binding protein		"nucleus"	Nucleus
DKFZp586J1023	AL136938	1048	72	749	226	protein management	similar to: glutathione S-transferase / posttranslational modification		"no predict"	Cytosol + Nucleus
DKFZp586J1119	AL136919	2343	28	2151	708	signaling & communication	unknown		"membranes"	Endoplasmic Reticulum
DKFZp586J1923	AL050220	745	49	588	179	differentiation &	similar to: serine protease	19	"secre pathway"	Endoplasmic Reticulum

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
						development				
DKFZp586K0919	AL50283	1782	204	1316	371	unknown	unknown		"no predict"	Cytosol + Nucleus
DKFZp586L0118	AL136913	1076	45	596	184	protein management	similar to: mitochondrial Ribosomal S40 protein		"nucleus"	Nucleus
DKFZp586M2420	AL136927	1986	23	1855	611	transport and traffic	similar to: mannosyltransferase	11	"secr pathway"	Endoplasmic Reticulum
DKFZp727E151	AL390215	1957	340	1701	454	membrane protein	similar to: transporter proteins (contains 9 transmembrane domains)		"no predict"	Endoplasmic Reticulum
DKFZp727M111	AL117479	2275	79	1899	633	unknown	unknown		"no predict"	Cytosol
DKFZp727M231	AL117480	2428	56	1581	542	unknown	unknown		"no predict"	Cytosol
DKFZp761G05121	AL118986	4592	107	3613	1169	protein management	similar to: SH3 BINDING PROTEIN		"cytosol"	Cytosol
DKFZp761G18121	AL136548	4117	107	1438	444	signaling & communication	similar to: ALLOGRAFT INFLAMMATORY FACTOR		"nucleus"	Nucleus
DKFZp761I12121	AL136549	4130	139	3894	1252	cell cycle	similar to: p53 inducible protein	5q34	"no predict"	Cytosol
DKFZp761M02121	AL136551	3328	178	2163	662	cell cycle	similar to: p53 regulated PA26-T2 nuclear protein		"nucleus"	Cytosol
DKFZp761O15121	AL136552	4293	112	2421	770	signaling & communication	similar to: semaphorin W	328.8 cR from top of Chr2 linkage group	"secr pathway"	Endoplasmic Reticulum

Table of cDNA clones and related data

Group: cell cycle

Cloned	Accession	Function	Group
hfr2_16g18	Similarity to KIAA0797 and yeast Smt4p	Novel protein with similarities to S. pombe SPAC17A5.07c and the S. cerevisiae Smt4p suppressor of Mif2 Gene.; involved in centromere organisation	Cell cycle
hfr2_2k14	Strong similarity to human N13 tumour suppressor gene	New tumour suppressor gene	Cell cycle
htes3_35b4	Human M-phase phosphoprotein-1	The novel protein is C-terminal identical to human M-phase phosphoprotein-1, which is expressed and phosphorylated in the metaphase. Therefore the novel protein seems to be involved in the mitotic spindle during cell division.	Cell cycle
htes3_35p22	Strong similarity to oncogene 1 (cre-2 locus)	Oncogene	Cell cycle
htes3_7j3	Related to the C-TAK1 Cdc25C associated protein kinase	Cdc25C is a protein kinase that controls entry into mitosis by dephosphorylation of Cdc2. Cdc25C function is regulated by phosphorylation, too. Serine 216 phosphorylation of Cdc25C mediates the binding of 14-3-3 protein to Cdc25C. C-TAK1 (Cdc twenty-five	Cell cycle
htes3_7p10	Strong similarity to XPMC2 protein	XPMC2 of xenopus rescues several different yeast mitotic catastrophe mutants defective in Wee1/Mik1 kinase function.	Cell cycle
hute1_20m11	Similarity to suppressor protein sds22	Suppressor regulator of protein phosphatase-1	Cell cycle

Group cell structure and motility

Gene ID	Gene Name	Function	Group
hfbr2_16c16	Similarity to Drosophila kelch	Shares the features of mayven and kelch and therefore should be involved in the organisation of cyto skeleton binding to membrane proteins	Structure and motility
hfbr2_2b5	Similarity to collagen proteins	New collagen alpha chain	Structure and motility
htes3_15i5	Strong similarity to "radial spokehead" proteins	Part of sperm motor	Structure and motility
htes3_18l7	Similarity to ankyrins	Putative ankyrin	Structure and motility
htes3_1kl1	Strong similarity to mouse ENC-1	Nuclear matrix protein	Structure and motility
htes3_72kl5	Strong similarity to Rattus norvegicus actin-filament binding protein Frabin.	FGD1-related F-actin-binding protein (Frabin/FGD1) is a novel F-actin binding protein. Modulation of cell structure and motility as well as modulation of the JNK/SAPK pathway.	Structure and motility
htes3_7b22	Similarity to paramyosins	Protein involved in motility	Structure and motility
hutel_19g22	Strong similarity to tuftelin/enamelin	New connective tissue protein	Structure and motility
hutel_24j6	Strong similarity Rattus norvegicus cell adhesion regulator (CAR1) mRNA	Cell adhesion regulator (signal transduction molecule influencing cell adhesion to collagen)	Structure and motility

Group Differentiation/Development

Clone ID	Homology	Function	Group
hfr2_2d15	Mus musculus testis-specific Y-encoded-like protein (Tspyl1).	TSPY is believed to function in early spermatogenesis and is a candidate for GBY, the putative gonadoblastoma-inducing gene on the Y-chromosome	Differentiation/Development
htes3_35e21	Similarity to interleukin-7 precursor	New interleukin	Differentiation/Development
hute1_2h3	Strong similarity to mouse E25 and gallus E3-16	Homolog is marker for chondro-osteogenic differentiation	Differentiation/Development

Group kidney derived

CloneID	Phenology	Function	Group
hfk2_1j9	Strong similarity to XLCL2 protein, African clawed frog	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Kidney derived
hfk2_24e23	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Kidney derived
hfk2_46a6	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Kidney derived
hfk2_46b10	Similarity to C.elegans F25B5.3	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Kidney derived
hfk2_46d13	Weak similarity to KE03 protein	Contains a RGD site; No informative BLAST results; No predictive prosite, pfam or SCOP motive	Kidney derived
hfk2_4b6	Similarity to Homo sapiens clone 25003 partial CDS.	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Kidney derived
hfk2_4c8	Similarity to KIA0549 and HAP1 (Huntingtin-associated protein-1)	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Kidney derived

Group mammary carcinoma derived

Clonotype ID	Sequence	Function	Group
hmcfl_1c23	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Mammary Carcinoma derived
hmcfl_1g13	Similarity to KIAA0766; very weak similarity to transposases	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Mammary Carcinoma derived

Group Nucleic acid management

Gene/Protein	Accession	Function	Group
hfr2_23b10	Similarity to rat RNA helicase HEL17	RNA helicase	Nucleic Acid Management
hfr2_3cl8	Strong similarity to RNA helicase and RNA-dependent ATPase from the DEAD box family	DEAD-box	Nucleic Acid Management
hfr2_64a15	Similarity to inorganic pyrophosphatases (unapplied)	Inorganic pyrophosphatase	Nucleic Acid Management
hfr2_6ol7	Strong similar to RNA helicases	RNA helicases	Nucleic Acid Management
hfr2_72b18	Similarity to DNA damage induced genes	Similar to dinP of <i>E. coli</i> , YqjH of <i>B. subtilis</i> , dinP of <i>M. tuberculosis</i> and T19K24.15 of <i>A. thaliana</i> . The dinB/P pathway is a second SOS-pathway in <i>E. coli</i>	Nucleic acid management
hfr2_72l12	Similarity to YDR126w	DNA binding protein	Nucleic acid management
hfr2_82f24	Strong similarity to DEAD-box subfamily ATP-dependent helicase	Dead-box helicase	Nucleic Acid Management
htes3_14h21	Strong similarity to RNA helicases	RNA helicase	Nucleic Acid Management
htes3_15j3	Similarity to YGR276c, a ribonuclease H of <i>S. cerevisiae</i> .	Rnase H	Nucleic Acid Management
htes3_20m18	Similarity to the <i>S. cerevisiae</i> mitochondrial carrier protein RIM2.	The novel protein contains a leucine zipper and a Prosite mitochondrial energy transfer proteins signature. It is member of a family of substrate carrier proteins which are found in the inner mitochondrial membrane and are involved in energy transfer.	Nucleic Acid Management
htes3_22g2	KIAA0829 is shorter, nearly identical to rat TIP120	Involved in TATA box binding complex	Nucleic Acid Management
htes3_2m18	Nearly identical to mouse Dhml	Multifunctional nuclease/exoribonuclease	Nucleic acid management
htes3_7p9	Similarity to nuclear domain 10 protein NDP52	Transcription control	Nucleic Acid Management
htes3_8m10	Strong similarity to polyadenylate-binding proteins.	The poly(A)-binding protein (PABP) binds to the messenger (mRNA) 3'-poly(A) tail found on most eukaryotic mRNAs and together with the poly(A) tail has been implicated in governing the stability and the translation of mRNA.	Nucleic Acid Management
htes3_18l1	Strong similarity to <i>S. cerevisiae</i> YHR148w	Mitochondrial Ribosomal S40 protein	Nucleic Acid Management

Group testis associated

Accession	Protein	Function	Group
htes3_14g5	Strong similarity to cell growth regulating nucleolar protein LYAR, of mouse	Contains a ATP/GTP-binding site motif A (P-loop), but not the zinc finger motif and nuclear localization signals of lyar.	Testes associated
htes3_14p14	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_14p7	Weak similarity to kinesin associated protein KAP3	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_15a13	Similarity to S.cerevisiae Hop1	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_15g14	Similarity to YOR243c	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_15b1	Weak similarity to Hsp70/Hsp90 organizing protein	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_15j18	Unknown	Unknown; no predictive prosite pfam or SCOP motive	Testes associated
htes3_17f10	T2J87-2B PROTEIN	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_18f3	Similarity to TNF-inducible protein CG12-1	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_19f19	Weak similarity to S. cerevisiae protein YFL046w.	The protein contains a RGD cell attachment site.	Testes associated
htes3_19j17	Partial similarity to C.elegans Y4081A.2 protein.	No informative BLAST results; No predictive prosite, pfam or SCOP motive.	Testes associated
htes3_20c21	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive.	Testes associated
htes3_21n23	Strong similarity to rat 7a comp protein	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_22c23	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_22n13	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_27o14	Similarity to C.elegans C55A6.1	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_28d14	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_2a11	Similarity to mucin	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_2d15	Similarity to C.elegans P25H2.1	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_2f14	Weak similarity to omega protein	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_2g7	Similarity to neurofilament proteins	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_2h15	Similarity to S.pombe cdc23	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_2i19	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive.	Testes associated

Cloned ID	Technology	Function	Group
htes3_2m20	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive.	Testes associated
htes3_2n9	Very weak similarity to Homo sapiens PAC clone DJ0771P04 from 7q11.21-q11.23.	No informative BLAST results; No predictive prosite, pfam or SCOP motive.	Testes associated
htes3_30f4	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive.	Testes associated
htes3_35g6	Strong similarity to R27216_1	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_35n24	Unknown	Contains an IG_MHC pattern	Testes associated
htes3_35p17	Similarity to S.cerevisiae VACB and beta-Catenin, but contains no amadillo motifs	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_4b4	Rattus norvegicus late gestation lung protein 1	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_4f17	Similarity to KIAA0333 Methyl-CpG binding protein; does not contain such a motive.	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_4o19	Similarity to mucin	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_50j4	Unknown, prolin rich protein	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_50n23	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_50n6	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_6b21	Similarity to KIAA0256	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_6dl6	WUGSC:H DJ1185107.2, differences to genmodel	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_72kl1	Similarity to S.pombe hypothetical repeat-containing protein	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_7dl7	Similarity to KIAA0454	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_7j8	WUGSC:H DJ1159004.1 similarity to YBL104p	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_8g11	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_8g5	KIAA087, alternative spliced	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_8p7	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_9e22	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_9i20	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_9k22	Similarity to C-terminus of katanin p80	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated

Group transmembrane proteins

Accession Number	Protein	Function	Group
hfr2_16112	Similarity to Fugu rubripes POT2	1 transmembrane domain No informative BLAST results	Transmembran e protein
hfr2_16112	Similarity to gallus putative transmembrane protein E3-16	1 transmembrane domain No informative BLAST results	Transmembran e protein
hfr2_22h13	Similarity to Drosophila melanogaster EG:39E1.3.	1 transmembrane domain No informative BLAST results	Transmembran e protein
hfr2_2b17	Similarity to Drosophila hypothetical 30K protein	3 transmembrane domains No informative BLAST results	Transmembran e protein
hfr2_2d17	Unknown	1 transmembrane domain No informative BLAST results	Transmembran e protein
hfr2_64k24	Similarity to several proteins	5 transmembrane regions No informative BLAST results	Transmembran e protein
hfr2_82c20	Similarity to C.elegans D1007.5	7 transmembrane domains No informative BLAST results	Transmembran e protein
hfr2_82e17	Similarity to C.elegans "R01B10.5"	6 transmembrane domains No informative BLAST results	Transmembran e protein
hfr2_82g14	Unknown proline rich protein	1 transmembrane domain No informative BLAST results	Transmembran e protein
hfr2_24a15	Similarity to C. elegans R07G3.8	1 transmembrane domain No informative BLAST results	Transmembran e protein
hfr2_31i3	Similarity to A.thaliana YUP8H12.2	3 transmembrane domains No informative BLAST results	Transmembran e protein
hfr2_4m11	Weak similarity to YMR034c	4 transmembrane domains No informative BLAST results	Transmembran e protein
hmcfl_1a11	Similarity to YDR255c and SPBC29A3.03c	1 transmembrane domain No informative BLAST results	Transmembran e protein
hmcfl_1e15	Similarity to D-XYLOSE TRANSPORTER	Transporter: 9 transmembrane domains No informative BLAST results	Transmembran e protein
htes3_15c6	Unknown	1 transmembrane domain No informative BLAST results	Transmembran e protein
htes3_2ol3	Partial similarity to the IL-17 receptor.	1 transmembrane domain No informative BLAST results	Transmembran e protein
htes3_27k4	Strong similarity to C.elegans K07H8.2/ZK185.2	Contains a leucine zipper 10 transmembrane domains No informative BLAST results	Transmembran e protein
htes3_2h1	Similarity to C.elegans C13P10.5	1 transmembrane domain No informative BLAST results	Transmembran e protein
htes3_35k24	Unknown	5 transmembrane domains No informative BLAST results	Transmembran e protein
htel_19f19	Similarity to mouse P24 protein	2 transmembrane domains No informative BLAST results	Transmembran e protein
htel_24c19	Unknown	1 transmembrane domain No informative BLAST results	Transmembran e protein

Group Brain derived

Clonotype ID	Homology	Antigen	Group
hhfbr2_16f21	Strong similarity to zinc finger protein 216 has no zn finger, is only similar	PROSITE: Contains no Zinc finger; No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hhfbr2_16k22	Weak similarity to thioredoxin	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hhfbr2_22f21	Weak similarity to C.elegans C18C4.5	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hhfbr2_22i4	Similarity to Human P52ripk N-terminus	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hhfbr2_22k3	Weak homology with : EXTENSIN (PROLINE-RICH GLYCOPROTEIN)	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hhfbr2_22k8	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hhfbr2_23f2	Similarity to Vps29p; <i>saccharomyces cerevisiae</i> (baker's yeast) pep11 protein	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hhfbr2_23o24	Similarity to CAAX-box protein	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hhfbr2_23o5	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hhfbr2_2a2	Similarity to 52K autoantigen Ro/SS-A - human	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hhfbr2_2c1	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hhfbr2_2c18	Weak similarity to cyclin-dependent kinase p130-PITSLRE	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hhfbr2_2d20	Similarity to <i>Synechocystis</i> sp. (PCC 6803)	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hhfbr2_2g18	J30M3.2 extension of genmodel	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hhfbr2_2h1	Similarity to C.elegans D2007.4 protein	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hhfbr2_2h10	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hhfbr2_2k19	Similarity to KIAA0378	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hhfbr2_3f16	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hhfbr2_3l2	Weak similarity to ubiquitin-like protein DSK2 yeast	Pfam: ubiquitin family; No informative BLAST results; No predictive prosite or SCOP motive	Brain derived
hhfbr2_62n10	Similarity to reticulocyte-binding protein	Contains a Leucine zipper; No informative BLAST results; No predictive pfam or SCOP motive	Brain derived
hhfbr2_64a11	Similarity to <i>Drosophila</i> irregular chiasm C-roughest precursor (frame shift)	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived

Cloned protein	Homology	Function	Group
hfr2_64c16	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfr2_64c4	Similarity to A. thaliana T08113.5	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfr2_64h6	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfr2_64i20	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfr2_64o16	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfr2_6a17	Weak similarity to finger protein zfOC1	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfr2_6i20	Similarity to ribosomal protein L15 precursor, mitochondrial	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfr2_7i020	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfr2_72d13	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfr2_72m16	Similarity to C.elegans H14A12.3	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfr2_72n12	Strong similarity to rat Ganglioside expression factor (GEP-2) but even higher identity with C.elegans putative protein Identities = 91/116 (78%)	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfr2_78d13	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfr2_78n23	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfr2_7a24	DKF2phfr2_7a24.1 similarity to C- terminus of TGF-beta-activated kinase	Only c-terminus homolog; contains no kinase domain; No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfr2_7e22	Similarity to cytochrome b561	No heme domain but a c may helix loop helix signature	Brain derived
hfr2_7j4	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfr2_82m16	Very weak similarity to A.thaliana F28A23.140	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived

Group Intracellular Transport and Trafficking

Gene	Homology	Function	Group
hfr2_23124	Strong similarity to human GP16b glycoprotein and canine VIP 36	A lectin character is predicted. Due to the intracellular localisation of the homologue proteins, it should be involved in cell trafficking	Transport and traffic
hfr2_2117	Strong similarity to rab1	GTP binding	Transport and traffic
hfr2_41m15	Strong similarity to ras-related GTP-binding protein Rab17	GTP-binding, signal transduction	Transport and traffic
hfr2_62f10	Strong similarity to zinc transporter proteins	Zinc transporter protein	Transport and traffic
hfr2_62l19	2 nearly identical to dog GTP-binding protein rab10	GTP binding	Transport and traffic
hfr2_64j18	Strong similarity to dog signal peptidase (GC 3.4.99.-)	Identical to canin and chicken microsomal signal peptidase 23 kd subunit.	Transport and traffic
hfr2_24n20	Strong similarity to eps8 binding protein e381	Contains an Src homology domain 3 and is similar to human eps8 SH3 domain binding protein 1 (e381) and spectrins. The new protein seems to be part of the signalling pathway between tyrosine kinases and the membrane/cyto skeleton	Transport and traffic
hfr2_24p5	Human ankyrin G (ANK-3) new splice variant	New ankyrin protein	Transport and traffic
hfr2_4k14	Strong similarity to Rab6	New Rab protein	Transport and traffic
htes1_1g13	Similarity to 256 kD golgin, strong similarity to rat "Cp151"	New golgin protein	Transport and traffic
htes1_17n18	TonB-dependent receptor protein signature 1	Involved in receptor-mediated uptake	Transport and traffic
htes1_21l16	Identical to rat ribosome attached membrane protein 4	Responsible for transport of proteins into ER	Transport and traffic
htes1_21l11	Nearly identical to mouse ADP-ribosylation-like factor homolog 6 (Ar16).	Protein secretion through the endoplasmic reticulum and the Golgi vesicular trafficking system is initiated by the binding of ADP-ribosylation factors	Transport and traffic
htes1_26g22	Similarity to kinesins.	The novel protein contains a ATP/GTP-binding site motif A (P-loop) and a kinesin motor domain signature. Kinesin is a microtubule-associated force-producing protein that play a role in organelle transport.	Transport and traffic
htes1_4h6	Strong similarity to Kinesin light chain	New kinesin light chain	Transport and traffic
htes1_72p16	Strong similarity to mouse MEM3 and yeast VPS35	New vacuolar protein sorting-associated protein	Transport and traffic
hutel_19h17	Strong similarity to C.elegans ZK1086.1	Steroid turnover in cells	Transport and traffic
hutel_20h13	Strong similarity to alpha-adaptins	New adaptin chain (clathrin assembly protein complex 2 alpha-a large chain)	Transport and traffic
hutel_24e11	Similarity to golgi 4-transmembrane spanning transporter mtp	New golgi transmembrane spanning transporter	Transport and traffic

Group signal transduction

GeneID	Accession	Function	Group
hfb2_23b21	Nearly identical to bovine neurocalcin	Neurocalcin is a Ca(2+)-binding protein with 3 EF-hands. Homology with recoverin indicates involvement in Ca2+ dependent activation of guanylate cyclase.	Signal transduction
hfb2_23n16	Similarity to putative phosphatidylinositol-4-phosphate 5-kinase	Contains a WW domain which binds proteins with particular proline- motifs, [AP]-P- [AP]-Y, and thus resembles somewhat SH3 domains. This domain is frequently associated with other domains typical for proteins in signal transduction processes	Signal transduction
hfb2_2c17	(similarity to YMR131c and retinoblastoma-binding protein RbAp46)	The protein contains 1 WD-40 repeat, which is typical for the beta-transducin subunit of G-proteins.	signal transduction
hfb2_22b11	Putative GTPase-activating protein, related to human chimaerins	The new protein is expected to activate p21rac-related small GTPases	Signal transduction
hfb2_78c24	Strong similarity to guanylate-binding proteins (GBPs)	Modulating/blocking the response of cells to interferons.	Signal transduction
hfb2_82e4	Strong similarity to rat calmodulin-binding protein	Involved in calmodulin-related pathway	Signal transduction
hfb2_82i17	Similarity to plasma membrane substrate for cAMP-dependent protein kinase	Transmembrane protein with strong similarity to the phospholipase protein, a membrane substrate for the cAMP-dependent protein kinase; seems to serve as chloride channels or as chloride-channel regulators. Transmembrane Protein	Signal Transduction
hfb2_82m6	Strong similarity to mouse "sphingosine kinase"	Sphingosine kinase	Signal transduction
hfd2_46m4	Nearly identical to mouse GTP-binding protein	GTP-binding protein	Signal transduction
htes3_15k11	KIAA0781, 5' extension	Heart development/signal transduction	Signal transduction
htes3_1c1	Similarity to GTPase-activating proteins	GTPase-activating proteins	Signal transduction
htes3_1n3	Similarity to Tup1p	Beta-transducin subunit of G-proteins	Signal transduction
htes3_20k2	Strong similarity to rat vanilloid receptor subtype 1.	VR1 seems to play an important role in the activation and sensitization of nociceptors. It is the receptor for e.g. capsaicin, a selective activator of nociceptors, a natural product of capsicum peppers. The novel protein is the human orthologue of rat VR1.	Signal transduction
htes3_21d4	Similarity to RCC1-like G exchanging factor RLK	RCC1 is a eukaryotic protein which binds to chromatin and interacts with ran, a nuclear GTP-binding protein.	Signal transduction
htes3_23n19	Similarity to rat protein kinase C-interacting RBCC protein 1	Protein kinase C-interacting protein	Signal transduction
htes3_4f5	Similarity to S.pombe "beta-transducin"	Contains 3 WD-40 repeats, which are typical for the beta-transducin subunit of G-proteins; in addition, a Cytochrome C family heme-binding site signature is present.	Signal Transduction
htes3_6c11	Strong similarity to YNL132w	Could be a steroid receptor	Signal transduction
htes3_8e24	Related to yeast YGL099w and mouse MMR1 putative GTP-binding proteins.	a novel 658 amino acid putative GTP-binding protein	Signal Transduction
hute1_20g21	Ras inhibitor	Receptor tyrosine kinase (RTK)/RAS/MAP kinase signaling cascade	Signal transduction
hute1_22d2	Similarity to GTP-binding proteins	GTP-binding proteins	Signal transduction

hutel_22e12	Strong similarity to <i>S.cerevisiae</i> YGL054c and cornichon	The <i>Drosophila</i> cni and mammalian proteins cornicon are part of a signal transduction pathway involving hte EGF-receptor	Signal transduction
-------------	--	--	---------------------

Group Metabolism

GeneID	Homology	Function	Group
hfr2_3g8	Similarity to N-terminal Acetyltransferase Complex ARD1 homolog	In yeast, ARD1 and NAT1, are required for the expression of an N-terminal protein acetyltransferase 1.	Metabolism
hfr2_62o17	Similarity to apolipoprotein E receptor	Low-density lipoprotein (LDL) receptors are the major cholesterol-carrying lipoproteins of plasma. The novel protein contains an additional leucine zipper suitable for protein-protein interaction.	Metabolism
hfr2_6b24	Similar to dTDP-6-deoxy-L-mannose-dehydrogenases	DTDP-6-deoxy-L-mannose-dehydrogenase	Metabolism
hfr2_78k24	Similarity to Mus musculus ubiquitin specific protease UBP43.	The novel protein contains a Prosite ubiquitin carboxyl-terminal hydrolases family 2 signature 2. These enzymes are involved in the processing of poly-ubiquitin precursors as well as that of ubiquitinated protein	Metabolism
hfr2_24b15	Similarity to phosphomannomutases	Phosphoserine signature typical for Phosphoglucosyltransferase or Phosphomannomutase ; conversion of Hexose phosphates.	Metabolism
hfr2_3o17	Strong similarity NADH Oxidoreductase B22 subunit.	The new protein is the human orthologue of the bovine EC 1.6.5.3. chain CI-B22 and therefore part of the human respiratory chain.	Metabolism
hfr2_46j20	Strong similarity to 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase	The new protein seems to be the human 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase.	Metabolism
hfr2_15c24	Strong similarity to 2-hydroxyacid dehydrogenases	New human 2-hydroxyacid dehydrogenase	Metabolism
hfr2_17l17	Strong similarity to transketolases	Transketolase testis specific	Metabolism
hfr2_27d1	Similarity to ubiquitin-specific proteases	Protease	Metabolism
hfr2_2a17	Similar to thiol-proteases	Putative thiol-protease	Metabolism
hfr2_35b5	Strong similarity to bovine vacuolar ATPase (EC 3.6.1.-) chain A	ATPase	Metabolism
hfr2_35k16	Similarity to acyl-CoA synthetase	Acyl-CoA synthetase	Metabolism
hfr2_35n12	Strong similarity to ADP/ATP carrier proteins	Involved in mitochondrial energy metabolism	Metabolism
hfr2_35n9	Carboxylesterase, splice variant	Carboxylesterase	Metabolism
hutel_20b19	Similarity to sarcosine oxidases	Sarcosine oxidases	Metabolism
hutel_20m24	Strong similarity to <i>S.cerevisiae</i> Alg9p probable mannosyltransferase	Possible mannosyltransferase	Metabolism
hutel_23e13	Strong similarity to heat shock 27K proteins	Heat shock protein related new subtilase	Metabolism

Group transcription factors

CloneID Accession	Homology	Function	Group
hfk2_46k19	Strong similarity to pterin-4-alpha-carbinolamine dehydratase	DcoH is a bifunctional protein, complexed with bioplerin. It serves as dimerization cofactor of hepatocyte nuclear factor-1 and catalyzes the dehydration of the bioplerin cofactor of phenylalanine hydroxylase	Transcriptio n factor
hfk2_47a4	Similarity to zinc fingers	New putative transcription factor with one C2H2 zinc fingers.	Transcriptio n factor
htes3_2e12	Similarity to finger proteins	Transcription factor with three C2H2 zinc fingers. Additionally, a cytochrome C family heme-binding site signature is present in the protein	Transcriptio n factor
htes3_2lj15	3 strong similarity to "NY-CO-33"	Transcription factor	Transcriptio n factor
htes3_17n12	Nearly identical to mouse SOX-LZ	SOX-LZ, related to SRY and HMG-box-Proteins	Transcriptio n factor
hutel_18i19	Similarity to transcription factor SF3	The SREBP-2 protein is cleaved to release soluble NH2-terminal that enter the nucleus and activate genes encoding the low density lipoprotein receptor and enzymes of cholesterol synthesis; a lim domain; shows similarity to the common sunflower transcript1	Transcriptio n factor
hutel_1i2	Similarity to Dictostelium myosin heavy chain kinase	Zn-finger protein	Transcriptio n factor

Group uterus associated

CloneID	Homology	Function	Group
hutel_17k7	Similarity to HPBRII-4 MRNA	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Uterus associated
hutel_18c12	Similarity to candidate tumor suppressor p33ING1	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Uterus associated
hutel_18i4	Weak similarity to C.elegans D2085.2	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Uterus associated
hutel_19g19	Partial similarity to bovine elastin fragment	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Uterus associated
hutel_19j11	Strong similarity to KIAA0231, similarity to ras binding protein Sur8	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Uterus associated
hutel_22n2	Similar to F46P6.1	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Uterus associated
hutel_21d15	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Uterus associated
hutel_22o2	Similarity to S.pombe SPBC3E7.03C	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Uterus associated
hutel_23g11	Similarity to SPAC31G5.12c and Maf1p	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Uterus associated

Prosite Key

NAME: N-glycosylation site.
 CONSENSUS: N-{P}-{ST}-{P}.

NAME: Glycosaminoglycan attachment site.
 CONSENSUS: S-G-x-G.

NAME: Tyrosine sulfation site.

NAME: cAMP- and cGMP-dependent protein kinase phosphorylation site.
 CONSENSUS: [RK](2)-x-{ST}.

NAME: Protein kinase C phosphorylation site.
 CONSENSUS: [ST]-x-[RK].

NAME: Casein kinase II phosphorylation site.
 CONSENSUS: [ST]-x(2)-[DE].

NAME: Tyrosine kinase phosphorylation site.
 CONSENSUS: [RK]-x(2,3)-[DE]-x(2,3)-Y.

NAME: N-myristoylation site.
 CONSENSUS: G-{EDRKHPFYW}-x(2)-[STAGCN]-{P}.

NAME: Amidation site.
 CONSENSUS: x-G-[RK]-[RK].

NAME: Aspartic acid and asparagine hydroxylation site.
 CONSENSUS: C-x-[DN]-x(4)-[FY]-x-C-x-C.

NAME: Vitamin K-dependent carboxylation domain.
 CONSENSUS: x(12)-E-x(3)-E-x-C-x(6)-[DEN]-x-[LIVMFY]-x(9)-[FYW].

NAME: Phosphopantetheine attachment site.
 CONSENSUS: [DEQGSTALMKRH]-[LIVMFYSTAC]-[GNQ]-[LIVMFYAG]-[DNEKHS]-S-[LIVMST]-
 CONSENSUS: {PCFY}-[STAGCPQLIVMF]-[LIVMATN]-[DENQGTAKRHLN]-[LIVMWSTA]-[LIVGSTACR]-
 CONSENSUS: x(2)-[LIVMFA].

NAME: Acyl carrier protein phosphopantetheine domain profile.

NAME: Prokaryotic membrane lipoprotein lipid attachment site.
 CONSENSUS: {DERK}(6)-[LIVMFWSTAG](2)-[LIVMFYSTAGCQ]-[AGS]-C.

NAME: Prokaryotic N-terminal methylation site.
 CONSENSUS: [KRHEQSTAG]-G-[FYLIVM]-[ST]-[LT]-[LIVP]-E-[LIVMFWSTAG](14).

NAME: Prenyl group binding site (CAAX box).
 CONSENSUS: C-{DENQ}-[LIVM]-x>.

NAME: Protein splicing signature.
 CONSENSUS: [DNEG]-x-[LIVFA]-[LIVMY]-[LVAST]-H-N-[STC].

NAME: Endoplasmic reticulum targeting sequence.
 CONSENSUS: [KRHQSA]-[DENQ]-E-L>.

NAME: Microbodies C-terminal targeting signal.
 CONSENSUS: [STAGCN]-[RKH]-[LIVMAFY]>.

NAME: Gram-positive cocci surface proteins 'anchoring' hexapeptide.
 CONSENSUS: L-P-x-T-G-[STGAVDE].

NAME: Bipartite nuclear targeting sequence.

NAME: Cell attachment sequence.
 CONSENSUS: R-G-D.

NAME: ATP/GTP-binding site motif A (P-loop).
 CONSENSUS: [AG]-x(4)-G-K-[ST].

NAME: Cyclic nucleotide-binding domain signature 1.
 CONSENSUS: [LIVM]-[VIC]-x(2)-G-[DENQTA]-x-[GAC]-x(2)-[LIVMFY](4)-x(2)-G.

NAME: Cyclic nucleotide-binding domain signature 2.

CONSENSUS: [LIVMF]-G-E-x-[GAS]-[LIVM]-x(5,11)-R-[STAQ]-A-x-[LIVMA]-x-[STACV].

NAME: cAMP/cGMP binding motif.

NAME: EF-hand calcium-binding domain.

CONSENSUS: D-x-[DNS]-{[ILVFW]}-[DENSTG]-[DNQGRK]-{GP}-[LIVMC]-[DENQSTAGC]-x(2)-

CONSENSUS: [DE]-[LIVMFYW].

NAME: Actinin-type actin-binding domain signature 1.

CONSENSUS: [EQ]-x(2)-[ATV]-[FY]-x(2)-W-x-N.

NAME: Actinin-type actin-binding domain signature 2.

CONSENSUS: [LIVM]-x-[SGN]-[LIVM]-[DAGHE]-[SAG]-x-[DNEAG]-[LIVM]-x-[DEAG]-x(4)-

CONSENSUS: [LIVM]-x-[LM]-[SAG]-[LIVM]-[LIVMT]-W-x-[LIVM](2).

NAME: Anaphylatoxin domain signature.

CONSENSUS: [CSH]-C-x(2)-[GAP]-x(7,8)-[GASTDEQR]-C-[GASTDEQL]-x(3,9)-[GASTDEQN]-x(2)-

CONSENSUS: [CE]-x(6,7)-C-C.

NAME: Anaphylatoxin domain profile.

NAME: Apple domain.

CONSENSUS: C-x(3)-[LIVMFY]-x(5)-[LIVMFY]-x(3)-[DENQ]-[LIVMFY]-x(10)-C-x(3)-C-T-

CONSENSUS: x(4)-C-x-[LIVMFY]-F-x-[FY]-x(13,14)-C-x-[LIVMFY]-[RK]-x-[ST]-x(14,15)-

CONSENSUS: S-G-x-[ST]-[LIVMFY]-x(2)-C.

NAME: Band 4.1 family domain signature 1.

CONSENSUS: W-[LIV]-x(3)-[KRQ]-x-[LIVM]-x(2)-[QH]-x(0,2)-[LIVMF]-x(6,8)-[LIVMF]-

CONSENSUS: x(3,5)-F-[FY]-x(2)-[DENS].

NAME: Band 4.1 family domain signature 2.

CONSENSUS: [HYW]-x(9)-[DENQSTV]-[SA]-x(3)-[FY]-[LIVM]-x(2)-[ACV]-x(2)-[LM]-x(2)-

CONSENSUS: [FY]-G-x-[DENQST]-[LIVMFYS].

NAME: Band 4.1 family domain profile.

NAME: C1q domain signature.

CONSENSUS: F-x(5)-[ND]-x(4)-[FYWL]-x(6)-F-x(5)-G-x-Y-x-F-x-[FY].

NAME: C-terminal cystine knot signature.

CONSENSUS: C-C-x(13)-C-x(2)-[GN]-x(12)-C-x-C-x(2,4)-C.

NAME: C-terminal cystine knot profile.

NAME: CUB domain profile.

NAME: Death domain profile.

NAME: EGF-like domain signature 1.

CONSENSUS: C-x-C-x(5)-G-x(2)-C.

NAME: EGF-like domain signature 2.

CONSENSUS: C-x-C-x(2)-[GP]-[FYW]-x(4,8)-C.

NAME: Calcium-binding EGF-like domain pattern signature.

CONSENSUS: [DEQN]-x-[DEQN](2)-C-x(3,14)-C-x(3,7)-C-x-[DN]-x(4)-[FY]-x-C.

NAME: Laminin-type EGF-like (LE) domain signature.

CONSENSUS: C-x(1,2)-C-x(5)-G-x(2)-C-x(2)-C-x(3,4)-[FYW]-x(3,15)-C.

NAME: Coagulation factors 5/8 type C domain (FA58C) signature 1.

CONSENSUS: [GAS]-W-x(7,15)-[FYW]-[LIV]-x-[LIVFA]-[GSTDEN]-x(6)-[LIVF]-x(2)-[IV]-x-

CONSENSUS: [LIVT]-[QKM]-G.

NAME: Coagulation factors 5/8 type C domain (FA58C) signature 2.

CONSENSUS: P-x(8,10)-[LM]-R-x-[GE]-[LIVP]-x-G-C.

NAME: Forkhead-associated (FHA) domain profile.

NAME: Fibrinogen beta and gamma chains C-terminal domain signature.

CONSENSUS: W-W-[LIVMFYW]-x(2)-C-x(2)-[GSA]-x(2)-N-G.

NAME: Type I fibronectin domain.

CONSENSUS: C-x(6,8)-[LFY]-x(5)-[FYW]-x-[RK]-x(8,10)-C-x-C-x(6,9)-C.

NAME: Type II fibronectin collagen-binding domain.

CONSENSUS: C-x(2)-P-F-x-[FYWI]-x(7)-C-x(8,10)-W-C-x(4)-[DNSR]-[FYW]-x(3,5)-[FYW]-x-

CONSENSUS: [FYWI]-C.

NAME: Hemopexin domain signature.

CONSENSUS: [LIFAT]-x(3)-W-x(2,3)-[PE]-x(2)-[LIVMFY]-[DENQS]-[STA]-[AV]-[LIVMFY].

NAME: Kringle domain signature.

CONSENSUS: [FY]-C-R-N-P-[DNR].

NAME: Kringle domain profile.

NAME: LDL-receptor class A (LDLRA) domain signature.

CONSENSUS: C-[VILMA]-x(5)-C-[DNH]-x(3)-[DENQHT]-C-x(3,4)-[STADE]-[DEH]-[DE]-x(1,5)-

CONSENSUS: C.

NAME: LDL-receptor class A (LDLRA) domain profile.

NAME: C-type lectin domain signature.

CONSENSUS: C-[LIVMFYATG]-x(5,12)-[WL]-x-[DNSR]-x(2)-C-x(5,6)-[FYWLIVSTA]-[LIVMSTA]-

CONSENSUS: C.

NAME: C-type lectin domain profile.

NAME: Link domain signature.

CONSENSUS: C-x(15)-A-x(3,4)-G-x(3)-C-x(2)-G-x(8,9)-P-x(7)-C.

NAME: Osteonectin domain signature 1.

CONSENSUS: C-x-[DN]-x(2)-C-x(2)-G-[KRH]-x-C-x(6,7)-P-x-C-x-C-x(3,5)-C-P.

NAME: Osteonectin domain signature 2.

CONSENSUS: F-P-x-R-[IM]-x-D-W-L-x-[NQ].

NAME: Somatomedin B domain signature.

CONSENSUS: C-x-C-x(3)-C-x(5)-C-C-x-[DN]-[FY]-x(3)-C.

NAME: Thyroglobulin type-1 repeat signature.

CONSENSUS: [FYWHP]-x-P-x-C-x(3,4)-G-x-[FYW]-x(3)-Q-C-x(4,10)-C-[FYW]-C-V-x(3,4)-

CONSENSUS: [SG].

NAME: P-type 'Trefoil' domain signature.

CONSENSUS: R-x(2)-C-x-[FYPST]-x(3,4)-[ST]-x(3)-C-x(4)-C-C-[FYWH].

NAME: Cellulose-binding domain, bacterial type.

CONSENSUS: W-N-[STAGR]-[STDN]-[LIVM]-x(2)-[GST]-x-[GST]-x(2)-[LIVMFT]-[GA].

NAME: Cellulose-binding domain, fungal type.

CONSENSUS: C-G-G-x(4,7)-G-x(3)-C-x(5)-C-x(3,5)-[NHG]-x-[FYWM]-x(2)-Q-C.

NAME: Chitin recognition or binding domain signature.

CONSENSUS: C-x(4,5)-C-C-S-x(2)-G-x-C-G-x(4)-[FYW]-C.

NAME: Barwin domain signature 1.

CONSENSUS: C-G-[KR]-C-L-x-V-x-N.

NAME: Barwin domain signature 2.

CONSENSUS: V-[DN]-Y-[EQ]-F-V-[DN]-C.

NAME: BIR repeat.

CONSENSUS: [HKEPILVY]-x(2)-R-x(3,7)-[FYW]-x(11,14)-[STAN]-G-[LMF]-X-[FYHDA]-X(4)-

CONSENSUS: [DESL]-X(2,3)-C-X(2)-C-X(6)-[WA]-X(9)-H-X(4)-[PRSD]-X-C-X(2)-[LIVMA].

NAME: WAP-type 'four-disulfide core' domain signature.

CONSENSUS: C-x-[C]-[DN]-x(2)-C-x(5)-C-C.

NAME: Phorbol esters / diacylglycerol binding domain.

CONSENSUS: H-x-[LIVMFYW]-x(8,11)-C-x(2)-C-x(3)-[LIVMFC]-x(5,10)-C-x(2)-C-x(4)-[HD]-

CONSENSUS: x(2)-C-x(5,9)-C.

NAME: C2 domain signature.

CONSENSUS: [ACG]-x(2)-L-x(2,3)-D-x(1,2)-[NGSTLIF]-[GTMR]-x-[STAP]-D-[PA]-[FY].

NAME: C2-domain profile.

NAME: CAP-Gly domain signature.

CONSENSUS: G-x(8,10)-[FYW]-x-G-[LIVM]-x-[LIVMFY]-x(4)-G-K-[NH]-x-G-[STAR]-x(2)-G-x(2)-[LY]-F.

NAME: Ly-6 / u-PAR domain signature.

CONSENSUS: [EQR]-C-[LIVMFYAH]-x-C-x(5,8)-C-x(3,8)-[EDNQSTV]-C-{C}-x(5)-C-x(12,24)-C.

NAME: MAM domain signature.

CONSENSUS: G-x-[LIVMFY](2)-x(3)-[STA]-x(10,11)-[LV]-x(4)-[LIVMF]-x(6,7)-C-[LIVM]-x-F-x-[LIVMFY]-x(3)-[GSC].

NAME: MAM domain profile.

NAME: PH domain profile.

NAME: Phosphotyrosine interaction domain (PID) profile.

NAME: Src homology 2 (SH2) domain profile.

NAME: Src homology 3 (SH3) domain profile.

NAME: VWFC domain signature.

CONSENSUS: C-x(2,3)-C-x-C-x(6,14)-C-x(3,4)-C-x(2,10)-C-x(9,16)-C-C-x(2,4)-C.

NAME: WW/rsp5/WWP domain signature.

CONSENSUS: W-x(9,11)-[VFY]-[FYW]-x(6,7)-[GSTNE]-[GSTQCR]-[FYW]-x(2)-P.

NAME: WW/rsp5/WWP domain profile.

NAME: ZP domain signature.

CONSENSUS: [LIVMFYW]-x(7)-[STAPDNL]-x(3)-[LIVMFYW]-x-[LIVMFYW]-x-[LIVMFYW]-x(2)-C-
CONSENSUS: [LIVMFYW]-x-[ST]-[PSL]-x(2,4)-[DENS]-x-[STADNQLF]-x(6)-[LIVM](2)-x(3,4)-
CONSENSUS: C.

NAME: S-layer homology domain signature.

CONSENSUS: [LVFYT]-x-[DA]-x(2,5)-[DNGSATPHY]-[WYFPDA]-x(4)-[LIV]-x(2)-[GTALV]-
CONSENSUS: x(4,6)-[LIVFYC]-x(2)-G-x-[PGSTA]-x(2,3)-[MFYA]-x-[PGAV]-x(3,10)-[LIVMA]-
CONSENSUS: [STKR]-[RY]-x-[EQ]-x-[STALIVM].

NAME: 'Homeobox' domain signature.

CONSENSUS: [LIVMFYG]-[ASLVR]-x(2)-[LIVMSTACN]-x-[LIVM]-x(4)-[LIV]-[RKNQUESTAIY]-
CONSENSUS: [LIVFSTNKH]-W-[FYVC]-x-[NDQTAH]-x(5)-[RKNAIMW].

NAME: 'Homeobox' domain profile.

NAME: 'Homeobox' antennapedia-type protein signature.

CONSENSUS: [LIVMFE]-[FY]-P-W-M-[KRQTA].

NAME: 'Homeobox' engrailed-type protein signature.

CONSENSUS: L-M-A-Q-G-L-Y-N.

NAME: 'Paired box' domain signature.

CONSENSUS: R-P-C-x(11)-C-V-S.

NAME: 'POU' domain signature 1.

CONSENSUS: [RKQ]-R-[LIM]-x-[LF]-G-[LIVMFY]-x-Q-x-[DNQ]-V-G.

NAME: 'POU' domain signature 2.

CONSENSUS: S-Q-[ST]-[TA]-I-[SC]-R-F-E-x-[LSQ]-x-[LI]-[ST].

NAME: Zinc finger, C2H2 type, domain.

CONSENSUS: C-x(2,4)-C-x(3)-[LIVMFYWC]-x(8)-H-x(3,5)-H.

NAME: Zinc finger, C3HC4 type (RING finger), signature.

CONSENSUS: C-x-H-x-[LIVMFY]-C-x(2)-C-[LIVMYA].

NAME: Nuclear hormones receptors DNA-binding region signature.

CONSENSUS: C-x(2)-C-x-[DE]-x(5)-[HN]-[FY]-x(4)-C-x(2)-C-x(2)-F-F-x-R.

NAME: GATA-type zinc finger domain.

CONSENSUS: C-x-[DN]-C-x(4,5)-[ST]-x(2)-W-[HR]-[RK]-x(3)-[GN]-x(3,4)-C-N-[AS]-C.

NAME: Poly(ADP-ribose) polymerase zinc finger domain signature.

CONSENSUS: C-[KR]-x-C-x(3)-I-x-K-x(3)-[RG]-x(16,18)-W-[FYH]-H-x(2)-C.

NAME: Poly(ADP-ribose) polymerase zinc finger domain profile.

NAME: Fungal Zn(2)-Cys(6) binuclear cluster domain signature.

CONSENSUS: [GASTPV]-C-x(2)-C-[RKHSTACW]-x(2)-[RKHQ]-x(2)-C-x(5,12)-C-x(2)-C-x(6,8)-

CONSENSUS: C.

NAME: Fungal Zn(2)-Cys(6) binuclear cluster domain profile.

NAME: Prokaryotic dksA/traR C4-type zinc finger.

CONSENSUS: C-[DES]-x-C-x(3)-I-x(3)-R-x(4)-P-x(4)-C-x(2)-C.

NAME: Copper-fist domain signature.

CONSENSUS: M-[LIVMF](3)-x(3)-K-[MY]-A-C-x(2)-C-I-[KR]-x-H-[KR]-x(3)-C-x-H-x(8)-

CONSENSUS: [KR]-x-[KR]-G-R-P.

NAME: Copper fist DNA binding domain profile.

NAME: Leucine zipper pattern.

CONSENSUS: L-x(6)-L-x(6)-L-x(6)-L.

NAME: bZIP transcription factors basic domain signature.

CONSENSUS: [KR]-x(1,3)-[RKSAQ]-N-x(2)-[SAQ](2)-x-[RKTAENQ]-x-R-x-[RK].

NAME: Myb DNA-binding domain repeat signature 1.

CONSENSUS: W-[ST]-x(2)-E-[DE]-x(2)-[LIV].

NAME: Myb DNA-binding domain repeat signature 2.

CONSENSUS: W-x(2)-[LI]-[SAG]-x(4,5)-R-x(8)-[YW]-x(3)-[LIVM].

NAME: Myc-type, 'helix-loop-helix' dimerization domain signature.

CONSENSUS: [DENSTAP]-K-[LIVMWAGSN]-{FYWCPHKR}-[LIVT]-[LIV]-x(2)-[STAV]-[LIVMSTAC]-x-

CONSENSUS: [VMFYH]-[LIVMTA]-{P}-{P}-[LIVMSR].

NAME: p53 tumor antigen signature.

CONSENSUS: M-C-N-S-S-C-M-G-G-M-N-R-R.

NAME: CBF-A/NF- κ B subunit signature.

CONSENSUS: C-V-S-E-x-I-S-F-[LIVM]-T-[SG]-E-A-[SC]-[DE]-[KRQ]-C.

NAME: CBF-B/NF- κ B subunit signature.

CONSENSUS: Y-V-N-A-K-Q-Y-x-R-I-L-K-R-R-x-A-R-A-K-L-E.

NAME: 'Cold-shock' DNA-binding domain signature.

CONSENSUS: [FY]-G-F-I-x(6,7)-[DER]-[LIVM]-F-x-H-x-[STKR]-x-[LIVMFY].

NAME: CTF/NF- κ B signature.

CONSENSUS: R-K-R-K-Y-F-K-K-H-E-K-R.

NAME: Ets-domain signature 1.

CONSENSUS: L-[FYW]-[QEDH]-F-[LI]-[LVQK]-x-[LI]-L.

NAME: Ets-domain signature 2.

CONSENSUS: [RKH]-x(2)-M-x-Y-[DENQ]-x-[LIVM]-[STAG]-R-[STAG]-[LI]-R-x-Y.

NAME: Ets-domain profile.

NAME: Fork head domain signature 1.

CONSENSUS: [KR]-P-[PTQ]-[FYLVQH]-S-[FY]-x(2)-[LIVM]-x(3,4)-[AC]-[LIM].

NAME: Fork head domain signature 2.

CONSENSUS: W-[QKR]-[NS]-S-[LIV]-R-H.

NAME: Fork head domain profile.

NAME: HSF-type DNA-binding domain signature.

CONSENSUS: L-x(3)-[FY]-K-H-x-N-x-[STAN]-S-F-[LIVM]-R-Q-L-[NH]-x-Y-x-[FYW]-[RKH]-K-

CONSENSUS: [LIVM].

NAME: Tryptophan pentad repeat (IRF family) signature.

CONSENSUS: W-x-[DNH]-x(5)-[LIVF]-x-[IV]-P-W-x-H-x(9,10)-[DE]-x(2)-[LIVF]-F-[KRQ]-x-

CONSENSUS: [WR]-A.

NAME: LIM domain signature.

CONSENSUS: C-x(2)-C-x(15,21)-[FYWH]-H-x(2)-[CH]-x(2)-C-x(2)-C-x(3)-[LIVMF].

NAME: LIM domain profile.

NAME: NF-kappa-B/Rel/dorsal domain signature.

CONSENSUS: F-R-Y-x-C-E-G.

NAME: MADS-box domain signature.

CONSENSUS: R-x-[RK]-x(5)-I-x-[DN]-x(3)-[KR]-x(2)-T-[FY]-x-[RK](3)-x(2)-[LIVM]-x-

CONSENSUS: K(2)-A-x-E-[LIVM]-[ST]-x-L-x(4)-[LIVM]-x-[LIVM](3)-x(6)-[LIVMF]-x(2)-

CONSENSUS: [FY].

NAME: MADS-box domain profile.

NAME: T-box domain signature 1.

CONSENSUS: L-W-x(2)-[FC]-x(3,4)-[NT]-E-M-[LIV](2)-T-x(2)-G-[RG]-[KRQ].

NAME: T-box domain signature 2.

CONSENSUS: [LIVMYW]-H-[PADH]-[DEN]-[GS]-x(3)-G-x(2)-W-M-x(3)-[IVA]-x-F.

NAME: TEA domain signature.

CONSENSUS: G-R-N-E-L-I-x(2)-Y-I-x(3)-[TC]-x(3)-R-T-[RK](2)-Q-[LIVM]-S-S-H-[LIVM]-

CONSENSUS: Q-V.

NAME: Transcription factor TFIIIB repeat signature.

CONSENSUS: G-[KR]-x(3)-[STAGN]-x-[LIVMYA]-[GSTA](2)-[CSAV]-[LIVM]-[LIVMFY]-[LIVMA]-

CONSENSUS: [GSA]-[STAC].

NAME: Transcription factor TFIIID repeat signature.

CONSENSUS: Y-x-P-x(2)-[IF]-x(2)-[LIVM](2)-x-[KRH]-x(3)-P-[RKQ]-x(3)-L-[LIVM]-F-x-

CONSENSUS: [STN]-G-[KR]-[LIVM]-x(3)-G-[TAGL]-[KR]-x(7)-[AGC]-x(7)-[LIVM].

NAME: TFIIIS zinc ribbon domain signature.

CONSENSUS: C-x(2)-C-x(9)-[LIVMQSAR]-[QH]-[STQL]-[RA]-[SACR]-x-[DE]-[DET]-[PGSEA]-

CONSENSUS: x(6)-C-x(2,5)-C-x(3)-[FW].

NAME: TSC-22 / dip / bun family signature.

CONSENSUS: M-D-L-V-K-x-H-L-x(2)-A-V-R-E-E-V-E.

NAME: Prokaryotic transcription elongation factors signature 1.

CONSENSUS: [ST]-x(2)-[GS]-x(3)-[LI]-x(2)-E-L-x(2)-L-x(3,4)-R-x(2)-[IV]-x(3)-[LIV]-

CONSENSUS: x(6)-G-D-x(2)-E-N-[GSA]-x-Y.

NAME: Prokaryotic transcription elongation factors signature 2.

CONSENSUS: S-x(2)-S-P-[LIVM]-[AG]-x-[SAG]-[LIVM]-[LIVMY]-x(4)-[DG]-[DE].

NAME: DEAD-box subfamily ATP-dependent helicases signature.

CONSENSUS: [LIVMF](2)-D-E-A-D-[RKEN]-x-[LIVMFYGSTN].

NAME: DEAH-box subfamily ATP-dependent helicases signature.

CONSENSUS: [GSAH]-x-[LIVMF](3)-D-E-[ALIV]-H-[NECR].

NAME: Eukaryotic putative RNA-binding region RNP-1 signature.

CONSENSUS: [RK]-G-[EDRKHPCG]-[AGSCI]-[FY]-[LIVA]-x-[FYLM].

NAME: Fibrillarin signature.

CONSENSUS: [GST]-[LIVMAP]-V-Y-A-[IV]-E-[FY]-[SA]-x-R-x(2)-R-[DE].

NAME: MCM family signature.

CONSENSUS: G-[IVT]-[LVAC](2)-[IVT]-D-[DE]-[FL]-[DNST].

NAME: MCM family domain.

NAME: XPA protein signature 1.

CONSENSUS: C-x-[DE]-C-x(3)-[LIVMF]-x(1,2)-D-x(2)-L-x(3)-F-x(4)-C-x(2)-C.

NAME: XPA protein signature 2.

CONSENSUS: [LIVM](2)-T-[KR]-T-E-x-K-x-[DE]-Y-[LIVMF](2)-x-D-x-[DE].

NAME: XPG protein signature 1.

CONSENSUS: [VI]-[KRE]-P-x-[FYIL]-V-F-D-G-x(2)-[PIL]-x-[LVC]-K.

NAME: XPG protein signature 2.

CONSENSUS: [GS]-[LIVM]-[PER]-[FYS]-[LIVM]-x-A-P-x-E-A-[DE]-[PAS]-[QS]-[CLM].

NAME: Bacterial regulatory proteins, araC family signature.

CONSENSUS: [KRQ]-[LIVMA]-x(2)-[GSTALIV]-{FYWPGDN}-x(2)-[LIVMSA]-x(4,9)-[LIVMF]-

CONSENSUS: x(2)-[LIVMSTA]-[GSTACIL]-x(3)-[GANQRF]-[LIVMFY]-x(4,5)-[LFY]-x(3)-

CONSENSUS: [FYIVA]-{FYWHCM}-x(3)-[GSADENQKR]-x-[NSTAPKL]-[PARL].

NAME: Bacterial regulatory proteins, araC family DNA-binding domain profile.

NAME: Bacterial regulatory proteins, arsR family signature.

CONSENSUS: C-x(2)-D-[LIVM]-x(6)-[ST]-x(4)-S-[HYR]-[HQ].

NAME: Bacterial regulatory proteins, asnC family signature.

CONSENSUS: [GSTAP]-x(2)-[DNEA]-[LIVM]-[GSA]-x(2)-[LIVMFY]-[GN]-[LIVMST]-[ST]-x(6)-R-

CONSENSUS: [LVT]-x(2)-[LIVM]-x(3)-G.

NAME: Bacterial regulatory proteins, crp family signature.

CONSENSUS: [LIVM]-[STAG]-[RHNW]-x(2)-[LIM]-[GA]-x-[LIVMFYA]-[LIVSC]-[GA]-x-[STACN]-

CONSENSUS: x(2)-[MST]-x-[GSTN]-R-x-[LIVMF]-x(2)-[LIVMF].

NAME: Bacterial regulatory proteins, deoR family signature.

CONSENSUS: R-x(3)-[LIVM]-x(3)-[LIVM]-x(16,17)-[STA]-x(2)-T-[LIVMA]-[RH]-[KRNA]-D-

CONSENSUS: [LIVMF].

NAME: Bacterial regulatory proteins, gntR family signature.

CONSENSUS: [LIVAPKR]-[PILV]-x-[EQTIVMR]-x(2)-[LIVM]-x(3)-[LIVMFYK]-x-[LIVFT]-

CONSENSUS: [DNGSTK]-[RGTLV]-x-[STAIVP]-[LIVA]-x(2)-[STAGV]-[LIVMFYH]-x(2)-[LMA].

NAME: Bacterial regulatory proteins, iclR family signature.

CONSENSUS: [GA]-x(3)-[DS]-x(2)-E-x(6)-[CSA]-[LIVM]-[GSA]-x(2)-[LIVM]-[FYH]-[DN].

NAME: Bacterial regulatory proteins, lacI family signature.

CONSENSUS: [LIVM]-x-[DE]-[LIVM]-A-x(2)-[STAGV]-x-V-[GSTP]-x(2)-[STAG]-[LIVMA]-x(2)-

CONSENSUS: [LIVMFYAN]-[LIVMC].

NAME: Bacterial regulatory proteins, luxR family signature.

CONSENSUS: [GDC]-x(2)-[NSTAVY]-x(2)-[IV]-[GSTA]-x(2)-[LIVMFYWCT]-x-[LIVMFYWCR]-x(3)-

CONSENSUS: [NST]-[LIVM]-x(5)-[NRHSA]-[LIVMSTA]-x(2)-[KR].

NAME: Bacterial regulatory proteins, lysR family signature.

CONSENSUS: [NQKRHSTAG]-[LIVMFYTA]-x(2)-[STAGLV]-[STAG]-x(4)-[LIVMYCTQR]-[PSTANLVER]-

CONSENSUS: x-[PSTAGQV]-[PSTAGNVMF]-[LIVMFA]-[STAGH]-x(2)-[LIVMF]-x(2)-[LIVMFW]-

CONSENSUS: [RKEAV]-x(2)-[LIVMFYNTAE]-x(3)-[LIMVT].

NAME: Bacterial regulatory proteins, marR family signature.

CONSENSUS: [STNA]-[LIA]-x-[RNGS]-x(4)-[LM]-[EIV]-x(2)-[GES]-[LFYW]-[LIVC]-x(7)-

CONSENSUS: [DN]-[RKQG]-[RK]-x(6)-T-x(2)-[GA].

NAME: Bacterial regulatory proteins, merR family signature.

CONSENSUS: [GSA]-x-[LIVMFA]-[ASM]-x(2)-[STACLIV]-[GSDENQR]-[LIVC]-[STANHK]-x(3)-

CONSENSUS: [LIVM]-[RHF]-x-[YW]-[DEQ]-x(2,3)-[GHDNQ]-[LIVMF](2).

NAME: Bacterial regulatory proteins, tetR family signature.

CONSENSUS: G-[LIVMFYS]-x(2,3)-[TS]-[LIVMT]-x(2)-[LIVM]-x(5)-[LIVQS]-[STAGENQH]-x-

CONSENSUS: [GPAR]-x-[LIVMF]-[FYST]-x-[HFY]-[FV]-x-[DNST]-K-x(2)-[LIVM].

NAME: Transcriptional antiterminators bglG family signature.

CONSENSUS: [ST]-x-H-x(2)-[FA](2)-[LIVM]-[EQK]-R-x(2)-[QNK].

NAME: Sigma-54 factors family signature 1.

CONSENSUS: P-[LIVM]-x-[LIVM]-x(2)-[LIVM]-A-x(2)-[LIVMF]-x(2)-[HS]-x-S-T-[LIVM]-S-R.

NAME: Sigma-54 factors family signature 2.

CONSENSUS: R-R-T-[IV]-[AT]-K-Y-R.

NAME: Sigma-54 factors family profile.

NAME: Sigma-70 factors family signature 1.

CONSENSUS: [DE]-[LIVMF](2)-[HEQS]-x-G-x-[LIVMFA]-G-L-[LIVMFYE]-x-[GSAM]-[LIVMAPI].

NAME: Sigma-70 factors family signature 2.

CONSENSUS: [STN]-x(2)-[DEQ]-[LIVM]-[GAS]-x(4)-[LIVMF]-[PSTG]-x(3)-[LIVMA]-x-[NQR]-

CONSENSUS: [LIVMA]-[EQH]-x(3)-[LIVFW]-x(2)-[LIVM].
NAME: Sigma-70 factors ECF subfamily signature.
CONSENSUS: [STAIV]-[PQDEL]-[DE]-[LIV]-[LIVTA]-Q-x-[STAV]-[LIVMFYC]-[LIVMAK]-x-
CONSENSUS: [GSTAIV]-[LIMFYWQ]-x(12,14)-[STAP]-[FYW]-[LIF]-x(2)-[IV].
NAME: Sigma-54 interaction domain ATP-binding region A signature.
CONSENSUS: [LIVMFY](3)-x-G-[DEQ]-[STE]-G-[STAV]-G-K-x(2)-[LIVMFY].
NAME: Sigma-54 interaction domain ATP-binding region B signature.
CONSENSUS: [GS]-x-[LIVMF]-x(2)-A-[DNEQASH]-[GNEK]-G-[STIM]-[LIVMFY](3)-[DE]-[EK]-
CONSENSUS: [LIVM].
NAME: Sigma-54 interaction domain C-terminal part signature.
CONSENSUS: [FYW]-P-[GS]-N-[LIVM]-R-[EQ]-L-x-[NHAT].
NAME: Sigma-54 interaction domain profile.
NAME: Single-strand binding protein family signature 1.
CONSENSUS: [LIVMF]-[NST]-[KRT]-[LIVM]-x-[LIVMF](2)-G-[NHRK]-[LIVM]-[GST]-x-[DET].
NAME: Single-strand binding protein family signature 2.
CONSENSUS: T-x-W-[HY]-[RNS]-[LIVM]-x-[LIVMF]-[FY]-[NGKR].
NAME: Bacterial histone-like DNA-binding proteins signature.
CONSENSUS: [GSK]-F-x(2)-[LIVMF]-x(4)-[RKEQA]-x(2)-[RST]-x-[GA]-x-[KN]-P-x-T.
NAME: Dps protein family signature 1.
CONSENSUS: H-[FW]-x-[LIVM]-x-G-x(5)-[LV]-H-x(3)-[DE].
NAME: Dps protein family signature 2.
CONSENSUS: [LIVMFY]-[DH]-x-[LIVM]-[GA]-E-R-x(3)-[LIF]-[GDN]-x(2)-[PA].
NAME: DNA repair protein radC family signature.
CONSENSUS: H-N-H-P-S-G.
NAME: recA signature.
CONSENSUS: A-L-[KR]-[IF]-[FY]-[STA]-[STAD]-[LIVMQ]-R.
NAME: RecF protein signature 1.
CONSENSUS: P-[ED]-x(3)-[LIVM](2)-x-G-[GSAD]-P-x(2)-R-R-x-[FY]-[LIVM]-D.
NAME: RecF protein signature 2.
CONSENSUS: [LIVMFY](2)-x-D-x(2,3)-[SA]-[EH]-L-D-x(2)-[KRH]-x(3)-L.
NAME: RecR protein signature.
CONSENSUS: C-x(2)-C-x(3)-[ST]-x(4)-C-x-I-C-x(4)-R.
NAME: Histone H2A signature.
CONSENSUS: [AC]-G-L-x-F-P-V.
NAME: Histone H2B signature.
CONSENSUS: [KR]-E-[LIVM]-[EQ]-T-x(2)-[KR]-x-[LIVM](2)-x-[PAG]-[DE]-L-x-[KR]-H-A-
CONSENSUS: [LIVM]-[STA]-E-G.
NAME: Histone H3 signature 1.
CONSENSUS: K-A-P-R-K-Q-L.
NAME: Histone H3 signature 2.
CONSENSUS: P-F-x-[RA]-L-[VA]-[KRQ]-[DEG]-[IV].
NAME: Histone H4 signature.
CONSENSUS: G-A-K-R-H.
NAME: HMG1/2 signature.
CONSENSUS: [FI]-S-[KR]-K-C-S-[EK]-R-W-K-T-M.
NAME: HMG-I and HMG-Y DNA-binding domain (A + T-hook).
CONSENSUS: [AT]-x(1,2)-[RK](2)-[GP]-R-G-R-P-[RK]-x.
NAME: HMG14 and HMG17 signature.
CONSENSUS: R-R-S-A-R-L-S-A-[RK]-P.
NAME: Bromodomain signature.

CONSENSUS: [STANVF]-x(2)-F-x(4)-[DNS]-x(5,7)-[DENQTF]-Y-[HFY]-x(2)-[LIVMFY]-x(3)-
 CONSENSUS: [LIVM]-x(4)-[LIVM]-x(6,8)-Y-x(12,13)-[LIVM]-x(2)-N-[SACF]-x(2)-[FY].

NAME: Bromodomain profile.

NAME: Chromo domain signature.

CONSENSUS: [FYI]-x-[LIVMC]-[KR]-W-x-[GDNR]-[FYWLE]-x(5,6)-[ST]-W-[ES]-[PSTDN]-x(3)-
 CONSENSUS: [LIVMC].

NAME: Chromo and chromo shadow domain profile.

NAME: Regulator of chromosome condensation (RCC1) signature 1.

CONSENSUS: G-x-N-D-x(2)-[AV]-L-G-R-x-T.

NAME: Regulator of chromosome condensation (RCC1) signature 2.

CONSENSUS: [LIVMFA]-[STAGC](2)-G-x(2)-H-[STAGLI]-[LIVMFA]-x-[LIVM].

NAME: Protamine P1 signature.

CONSENSUS: [AV]-R-[NFY]-R-x(2,3)-[ST]-x-S-x-S.

NAME: Nuclear transition protein 1 signature.

CONSENSUS: S-K-R-K-Y-R-K.

NAME: Nuclear transition protein 2 signature 1.

CONSENSUS: H-x(3)-H-S-[NS]-S-x-P-Q-S.

NAME: Nuclear transition protein 2 signature 2.

CONSENSUS: K-x-R-K-x(2)-E-G-K-x(2)-K-[KR]-K.

NAME: Ribosomal protein L1 signature.

CONSENSUS: [IM]-x(2)-[LIVA]-x(2,3)-[LIVM]-G-x(2)-[LMS]-[GSNH]-[PTKR]-[KRAV]-G-x-
 CONSENSUS: [LMF]-P-[DENSTK].

NAME: Ribosomal protein L2 signature.

CONSENSUS: P-x(2)-R-G-[STAIV](2)-x-N-[APK]-x-[DE].

NAME: Ribosomal protein L3 signature.

CONSENSUS: [FL]-x(6)-[DN]-x(2)-[AGS]-x-[ST]-x-G-[KRH]-G-x(2)-G-x(3)-R.

NAME: Ribosomal protein L5 signature.

CONSENSUS: [LIVM]-x(2)-[LIVM]-[STAC]-[GE]-[QV]-x(2)-[LIVMA]-x-[STC]-x-[STAG]-[KR]-
 CONSENSUS: x-[STA].

NAME: Ribosomal protein L6 signature 1.

CONSENSUS: [PS]-[DENS]-x-Y-K-[GA]-K-G-[LIVM].

NAME: Ribosomal protein L6 signature 2.

CONSENSUS: Q-x(3)-[LIVM]-x(2)-[KR]-x(2)-R-x-F-x-D-G-[LIVM]-Y-[LIVM]-x(2)-[KR].

NAME: Ribosomal protein L9 signature.

CONSENSUS: G-x(2)-[GN]-x(4)-V-x(2)-G-[FY]-x(2)-N-[FY]-L-x(5)-[GA]-x(3)-[STN].

NAME: Ribosomal protein L10 signature.

CONSENSUS: [DEH]-x(2)-[GS]-[LIVMF]-[STN]-[VA]-x-[DEQK]-[LIVMA]-x(2)-[LIM]-R.

NAME: Ribosomal protein L11 signature.

CONSENSUS: [RKN]-x-[LIVM]-x-G-[ST]-x(2)-[SNQ]-[LIVM]-G-x(2)-[LIVM]-x(0,1)-[DENG].

NAME: Ribosomal protein L13 signature.

CONSENSUS: [LIVM]-[KRV]-[GK]-M-[LIV]-[PS]-x(4,5)-[GS]-[NQEKRA]-x(5)-[LIVM]-x-[AIV]-
 CONSENSUS: [LFY]-x-[GDN].

NAME: Ribosomal protein L14 signature.

CONSENSUS: [GA]-[LIV](3)-x(9,10)-[DNS]-G-x(4)-[FY]-x(2)-[NT]-x(2)-V-[LIV].

NAME: Ribosomal protein L15 signature.

CONSENSUS: K-[LIVM](2)-[GAL]-x-[GT]-x-[LIVMA]-x(2,5)-[LIVM]-x-[LIVMF]-x(3,4)-
 CONSENSUS: [LIVMFC]-[ST]-x(2)-A-x(3)-[LIVM]-x(3)-G.

NAME: Ribosomal protein L16 signature 1.

CONSENSUS: [KR]-R-x-[GSAC]-[KQVA]-[LIVM]-W-[LIVM]-[KR]-[LIVM]-[LFY]-[AP].

NAME: Ribosomal protein L16 signature 2.

CONSENSUS: R-M-G-x-[GR]-K-G-x(4)-[FWKR].

NAME: Ribosomal protein L17 signature.
 CONSENSUS: I-x-[ST]-[GT]-x(2)-[KR]-x-K-x(6)-[DE]-x-[LIMV]-[LIVMT]-T-x-[STAG]-[KR].

NAME: Ribosomal protein L19 signature.
 CONSENSUS: [RT]-[KRSVY]-[GSA]-x-V-[RS]-[KR]-[SA]-K-L-Y-Y-L-R.

NAME: Ribosomal protein L20 signature.
 CONSENSUS: K-x(3)-[KRC]-x-[LIVM]-W-[IV]-[STNALV]-R-[LIVM]-N-x(3)-[RKH].

NAME: Ribosomal protein L21 signature.
 CONSENSUS: [IVT]-x(3)-[KR]-x(3)-[KRQ]-K-x(6)-G-[HF]-R-[RQ]-x(2)-T.

NAME: Ribosomal protein L22 signature.
 CONSENSUS: [RKQN]-x(4)-[RH]-[GAS]-x-G-[KRQS]-x(9)-[HDN]-[LIVM]-x-[LIVMS]-x-[LIVM].

NAME: Ribosomal protein L23 signature.
 CONSENSUS: [RK](2)-[AM]-[IVFYT]-[IV]-[RKT]-L-[STANQK]-x(7)-[LIVMFT].

NAME: Ribosomal protein L24 signature.
 CONSENSUS: [GDEN]-D-x-V-x-[IV]-[LIVMA]-x-G-x(2)-[KA]-[GN]-x(2,3)-[GA]-x-[IV].

NAME: Ribosomal protein L27 signature.
 CONSENSUS: G-x-[LIVM](2)-x-R-Q-R-G-x(5)-G.

NAME: Ribosomal protein L29 signature.
 CONSENSUS: [KNQS]-[PSTL]-x(2)-[LIMFA]-[KRGSAN]-x-[LIVYSTA]-[KR]-[KRH]-[DESTANRL]-
 CONSENSUS: [LIV]-A-[KRCQVT]-[LIVMA].

NAME: Ribosomal protein L30 signature.
 CONSENSUS: [IVT]-[LIVM]-x(2)-[LF]-x-[LI]-x-[KRHQEG]-x(2)-[STNQH]-x-[IVT]-
 CONSENSUS: x(10)-[LMS]-[LIV]-x(2)-[LIVA]-x(2)-[LMFY]-[IVT].

NAME: Ribosomal protein L31 signature.
 CONSENSUS: H-P-F-[FY]-[TI]-x(9)-G-R-[AV]-x-[KR].

NAME: Ribosomal protein L33 signature.
 CONSENSUS: Y-x-[ST]-x-[KR]-[NS]-x(4)-[PAT]-x(1,2)-[LIVM]-[EA]-x(2)-K-[FY]-[CSD].

NAME: Ribosomal protein L34 signature.
 CONSENSUS: K-[RG]-T-[FYWL]-[EQS]-x(5)-[KRHS]-x(4,5)-G-F-x(2)-R.

NAME: Ribosomal protein L35 signature.
 CONSENSUS: [LIVM]-K-[TV]-x(2)-[GSA]-[SAIL]-x-K-R-[LIVMFY]-[KRL].

NAME: Ribosomal protein L36 signature.
 CONSENSUS: C-x(2)-C-x(2)-[LIVM]-x-R-x(3)-[LIVMN]-x-[LIVM]-x-C-x(3,4)-[KR]-H-x-Q-x-Q.

NAME: Ribosomal protein L1e signature.
 CONSENSUS: N-x(3)-[KR]-x(2)-A-[LIVT]-x-S-A-[LIV]-x-A-[ST]-[SGA]-x(7)-[RK]-G-H.

NAME: Ribosomal protein L6e signature.
 CONSENSUS: N-x(2)-P-L-R-R-x(4)-[FY]-V-I-A-T-S-x-K.

NAME: Ribosomal protein L7Ae signature.
 CONSENSUS: [CA]-x(4)-[IV]-P-[FY]-x(2)-[LIVM]-x-[GSQ]-[KRQ]-x(2)-L-G.

NAME: Ribosomal protein L10e signature.
 CONSENSUS: R-x-A-[FYW]-G-K-[PA]-x-G-x(2)-A-R-V.

NAME: Ribosomal protein L13e signature.
 CONSENSUS: [KR]-Y-x(2)-K-[LIVM]-R-[STA]-G-[KR]-G-F-[ST]-L-x-E.

NAME: Ribosomal protein L15e signature.
 CONSENSUS: [DE]-[KR]-A-R-x-L-G-[FY]-x-[SAP]-x(2)-G-[LIVMFY](4)-R-x-R-V-x-R-G.

NAME: Ribosomal protein L18e signature.
 CONSENSUS: [KRE]-x-L-x(2)-[PS]-[KR]-x(2)-[RH]-[PSA]-x-[LIVM]-[NS]-[LIVM]-x-[RK]-
 CONSENSUS: [LIVM].

NAME: Ribosomal protein L19e signature.
 CONSENSUS: R-x-[KR]-x(5)-[KR]-x(3)-[KRH]-x(2)-G-x-G-x-R-x-G-x(3)-A-R-x(3)-[KQ]-
 CONSENSUS: x(2)-W-x(7)-R-x(2)-L-x(3)-R.

NAME: Ribosomal protein L21e signature.
 CONSENSUS: G-[DE]-x-V-x(10)-[GV]-x(2)-[FYH]-x(2)-[FY]-x-G-x-T-G.

NAME: Ribosomal protein L24e signature.
 CONSENSUS: [FY]-x-[GS]-x(2)-[IV]-x-P-G-x-G-x(2)-[FYV]-x-[KRHE]-x-D.

NAME: Ribosomal protein L27e signature.
 CONSENSUS: G-K-N-x-W-F-F-x-K-L-R-F>.

NAME: Ribosomal protein L30e signature 1.
 CONSENSUS: [STA]-x(5)-G-x-[QKR]-x(2)-[LIVM]-[KQT]-x(2)-[KR]-x-G-x(2)-K-x-[LIVM](3).

NAME: Ribosomal protein L30e signature 2.
 CONSENSUS: [DE]-L-G-[STA]-x(2)-G-[KR]-x(6)-[LIVM]-x-[LIVM]-x-[DEN]-x-G.

NAME: Ribosomal protein L31e signature.
 CONSENSUS: V-[KR]-[LIVM]-x(3)-[LIVM]-N-x-[AK]-x-W-x-[KR]-G.

NAME: Ribosomal protein L32e signature.
 CONSENSUS: F-x-R-x(4)-[KR]-x(2)-[KR]-[LIVM]-x(3)-W-R-[KR]-x(2)-G.

NAME: Ribosomal protein L34e signature.
 CONSENSUS: Y-x-[ST]-x-S-[NY]-x(5)-[KR]-T-P-G.

NAME: Ribosomal protein L35Ae signature.
 CONSENSUS: G-K-[LIVM]-x-R-x-H-G-x(2)-G-x-V-x-A-x-F-x(3)-[LI]-P.

NAME: Ribosomal protein L36e signature.
 CONSENSUS: P-Y-E-[KR]-R-x-[LIVM]-[DE]-[LIVM](2)-[KR].

NAME: Ribosomal protein L37e signature.
 CONSENSUS: G-T-x-[SA]-x-G-x-[KR]-x(3)-[ST]-x(0,1)-H-x(2)-C-x-R-C-G.

NAME: Ribosomal protein L39e signature.
 CONSENSUS: [KRA]-T-x(3)-[LIVM]-[KRQF]-x-[NHS]-x(3)-R-[NHY]-W-R-R.

NAME: Ribosomal protein L44e signature.
 CONSENSUS: K-x-[TV]-K-K-x(2)-L-[KR]-x(2)-C.

NAME: Ribosomal protein S2 signature 1.
 CONSENSUS: [LIVMFA]-x(2)-[LIVMFYC](2)-x-[STAC]-[GSTANQEK]-[STALV]-[HY]-[LIVMF]-G.

NAME: Ribosomal protein S2 signature 2.
 CONSENSUS: P-x(2)-[LIVMF](2)-[LIVMS]-x-[GDN]-x(3)-[DENL]-x(3)-[LIVM]-x-E-x(4)-
 CONSENSUS: [GNQKRH]-[LIVM]-[AP].

NAME: Ribosomal protein S3 signature.
 CONSENSUS: [GSTA]-[KR]-x(6)-G-x-[LIVMT]-x(2)-[NQSCH]-x(1,3)-[LIVFCA]-x(3)-[LIV]-
 CONSENSUS: [DENQ]-x(7)-[LMT]-x(2)-G-x(2)-G.

NAME: Ribosomal protein S4 signature.
 CONSENSUS: [LIVM]-[DE]-x-R-L-x(3)-[LIVMC]-[VMFYHQ]-[KRT]-x(3)-[STAGCF]-x-[ST]-x(3)-
 CONSENSUS: [SAI]-[KR]-x-[LIVMF](2).

NAME: Ribosomal protein S5 signature.
 CONSENSUS: G-[KRQ]-x(3)-[FY]-x-[ACV]-x(2)-[LIVMA]-[LIVM]-[AG]-[DN]-x(2)-G-x-
 CONSENSUS: [LIVM]-G-x-[SAG]-x(5,6)-[DEQ]-[LIVM]-x(2)-A-[LIVMF].

NAME: Ribosomal protein S6 signature.
 CONSENSUS: G-x-[KRC]-[DENQRH]-L-[SA]-Y-x-I-[KRNSA].

NAME: Ribosomal protein S7 signature.
 CONSENSUS: [DENSK]-x-[LIVMET]-x(3)-[LIVMFT](2)-x(6)-G-K-[KR]-x(5)-[LIVMF]-[LIVMFC]-
 CONSENSUS: x(2)-[STA].

NAME: Ribosomal protein S8 signature.
 CONSENSUS: [GE]-x(2)-[LIV](2)-[STY]-T-x(2)-G-[LIVM](2)-x(4)-[AG]-[KRHAYT].

NAME: Ribosomal protein S9 signature.
 CONSENSUS: G-G-G-x(2)-[GSA]-Q-x(2)-[SA]-x(3)-[GSA]-x-[GSTAV]-[KR]-[GSAL]-[LIF].

NAME: Ribosomal protein S10 signature.
 CONSENSUS: [AV]-x(3)-[GDNSR]-[LIVMSTA]-x(3)-G-P-[LIVM]-x-[LIVM]-P-T.

NAME: Ribosomal protein S11 signature.
 CONSENSUS: [LIVMF]-x-[GSTAC]-[LIVMF]-x(2)-[GSTAL]-x(0,1)-[GSN]-[LIVMF]-x-[LIVM]-
 CONSENSUS: x(4)-[DEN]-x-T-P-x-[PA]-[STCH]-[DN].

NAME: Ribosomal protein S12 signature.
 CONSENSUS: [RK]-x-P-N-S-[AR]-x-R.

NAME: Ribosomal protein S13 signature.
 CONSENSUS: [KRQS]-G-x-R-H-x(2)-[GSNH]-x(2)-[LIVMC]-R-G-Q.

NAME: Ribosomal protein S14 signature.
 CONSENSUS: [RP]-x(0,1)-C-x(11,12)-[LIVMF]-x-[LIVMF]-[SC]-[RG]-x(3)-[RN].

NAME: Ribosomal protein S15 signature.
 CONSENSUS: [LIVM]-x(2)-H-[LIVMFY]-x(5)-D-x(2)-[SAGN]-x(3)-[LF]-x(9)-[LIVM]-x(2)-
 CONSENSUS: [FY].

NAME: Ribosomal protein S16 signature.
 CONSENSUS: [LIVMT]-x-[LIVM]-[KR]-L-[STAK]-R-x-G-[AKR].

NAME: Ribosomal protein S17 signature.
 CONSENSUS: G-D-x-[LIV]-x-[LIVA]-x-[QEK]-x-[RK]-P-[LIV]-S.

NAME: Ribosomal protein S18 signature.
 CONSENSUS: [IV]-[DY]-Y-x(2)-[LIVMT]-x(2)-[LIVM]-x(2)-[FYT]-[LIVM]-[ST]-[DERP]-x-
 CONSENSUS: [GY]-K-[LIVM]-x(3)-R-[LIVMAS].

NAME: Ribosomal protein S19 signature.
 CONSENSUS: [STDNQ]-G-[KRQM]-x(6)-[LIVM]-x(4)-[LIVM]-[GSD]-x(2)-[LF]-[GAS]-[DE]-F-
 CONSENSUS: x(2)-[ST].

NAME: Ribosomal protein S21 signature.
 CONSENSUS: [DE]-x-A-[LY]-[KR]-R-F-K-[KR]-x(3)-[KR].

NAME: Ribosomal protein S3Ae signature.
 CONSENSUS: [LIV]-x-[GH]-R-[IV]-x-E-x-[SC]-L-x-D-L.

NAME: Ribosomal protein S4e signature.
 CONSENSUS: H-x-K-R-[LIVM]-[SAN]-x-P-x(2)-W-x-[LIVM]-x-[KR].

NAME: Ribosomal protein S6e signature.
 CONSENSUS: [LIVM]-[STAMR]-G-G-x-D-x(2)-G-x-P-M.

NAME: Ribosomal protein S7e signature.
 CONSENSUS: [KR]-L-x-R-E-L-E-K-K-F-[SAP]-x-[KR]-H.

NAME: Ribosomal protein S8e signature.
 CONSENSUS: R-x(2)-T-G-[GA]-x(5)-[HR]-K-[KR]-x-K-x-E-[LM]-G.

NAME: Ribosomal protein S12e signature.
 CONSENSUS: A-L-[KRQP]-x-V-L-x(2)-[SA]-x(3)-[DN]-G-L.

NAME: Ribosomal protein S17e signature.
 CONSENSUS: A-x-I-x-[ST]-K-x-L-R-N-[KR]-I-A-G-[FY]-x-T-H.

NAME: Ribosomal protein S19e signature.
 CONSENSUS: P-x(6)-[SAN]-x(2)-[LIVMA]-x-R-x-[ALIV]-[LV]-Q-x-L-[EQ].

NAME: Ribosomal protein S21e signature.
 CONSENSUS: L-Y-V-P-R-K-C-S-[SA].

NAME: Ribosomal protein S24e signature.
 CONSENSUS: [FA]-G-x(2)-[KR]-[STA]-x-G-[FY]-[GA]-x-[LIVM]-Y-[DN]-[SN].

NAME: Ribosomal protein S26e signature.
 CONSENSUS: [YH]-C-V-S-C-A-I-H.

NAME: Ribosomal protein S27e signature.
 CONSENSUS: [QK]-C-x(2)-C-x(6)-F-[GS]-x-[PSA]-x(5)-C-x(2)-C-[GS]-x(2)-L-x(2)-P-x-G.

NAME: Ribosomal protein S28e signature.
 CONSENSUS: E-[ST]-E-R-E-A-R-x-L.

NAME: DNA mismatch repair proteins mutL / hexB / PMS1 signature.

CONSENSUS: G-F-R-G-E-A-L.

NAME: DNA mismatch repair proteins mutS family signature.

CONSENSUS: [ST]-[LIVM]-x-[LIVM]-x-D-E-[LIVMY]-[GC]-[RKH]-G-[GST]-x(4)-G.

NAME: mutT domain signature.

CONSENSUS: G-x(5)-E-x(4)-[STAGC]-[LIVMAC]-x-R-E-[LIVMFT]-x-E-E.

NAME: DnaA protein signature.

CONSENSUS: I-[GA]-x(2)-[LIVMF]-[SGDNK]-x(0,1)-[KR]-x-H-[STP]-[STV]-[LIVM](2)-x-

CONSENSUS: [SA]-x(2)-[KRE]-[LIVM].

NAME: Small, acid-soluble spore proteins, alpha/beta type, signature 1.

CONSENSUS: K-x-E-[LIV]-A-x-[DE]-[LIVMF]-G-[LIVMF].

NAME: Small, acid-soluble spore proteins, alpha/beta type, signature 2.

CONSENSUS: [KR]-[SAQ]-x-G-x-V-G-G-x-[LIVM]-x-[KR](2)-[LIVM](2).

NAME: Zinc-containing alcohol dehydrogenases signature.

CONSENSUS: G-H-E-x(2)-G-x(5)-[GA]-x(2)-[IVSAC].

NAME: Quinone oxidoreductase / zeta-crystallin signature.

CONSENSUS: [GSD]-[DEQH]-x(2)-L-x(3)-[SA](2)-G-G-x-G-x(4)-Q-x(2)-[KR].

NAME: Iron-containing alcohol dehydrogenases signature 1.

CONSENSUS: [STALIV]-[LIVF]-x-[DE]-x(6,7)-P-x(4)-[ALIV]-x-[GST]-x(2)-D-[TAIVM]-

CONSENSUS: [LIVMF]-x(4)-E.

NAME: Iron-containing alcohol dehydrogenases signature 2.

CONSENSUS: [GSW]-x-[LIVTSACD]-[GH]-x(2)-[GSAE]-[GSHYQ]-x-[LIVTP]-[GAST]-[GAS]-x(3)-

CONSENSUS: [LIVMT]-x-[HNS]-[GA]-x-[GTAC].

NAME: Short-chain dehydrogenases/reductases family signature.

CONSENSUS: [LIVSPADNK]-x(12)-Y-[PSTAGNCV]-[STAGNQCIVM]-[STAGC]-K-{PC}-[SAGFR]-

CONSENSUS: [LIVMSTAGD]-x(2)-[LIVMFYW]-x(3)-[LIVMFYWGAPTHQ]-[GSACQRHM].

NAME: Aldo/keto reductase family signature 1.

CONSENSUS: G-[FY]-R-[HSAL]-[LIVMF]-D-[STAGC]-[AS]-x(5)-E-x(2)-[LIVM]-G.

NAME: Aldo/keto reductase family signature 2.

CONSENSUS: [LIVMFY]-x(9)-[KREQ]-x-[LIVM]-G-[LIVM]-[SC]-N-[FY].

NAME: Aldo/keto reductase family putative active site signature.

CONSENSUS: [LIVM]-[PAIV]-[KR]-[ST]-x(4)-R-x(2)-[GSTAEQK]-[NSL]-x(2)-[LIVMFA].

NAME: Homoserine dehydrogenase signature.

CONSENSUS: A-x(3)-G-[LIVMFY]-[STAG]-x(2,3)-[DNS]-P-x(2)-D-[LIVM]-x-G-x-D-x(3)-K.

NAME: NAD-dependent glycerol-3-phosphate dehydrogenase signature.

CONSENSUS: G-[AT]-[LIVM]-K-[DN]-[LIVM](2)-A-x-[GA]-x-G-[LIVMF]-x-[DE]-G-[LIVM]-x-

CONSENSUS: [LIVMFYW]-G-x-N.

NAME: FAD-dependent glycerol-3-phosphate dehydrogenase signature 1.

CONSENSUS: [IV]-G-G-G-x(2)-G-[STACV]-G-x-A-x-D-x(3)-R-G.

NAME: FAD-dependent glycerol-3-phosphate dehydrogenase signature 2.

CONSENSUS: G-G-K-x(2)-[GSTE]-Y-R-x(2)-A.

NAME: Mannitol dehydrogenases signature.

CONSENSUS: [LIVMY]-x-[FS]-x(2)-[STAGCV]-x-V-D-R-[IV]-x-[PS].

NAME: Histidinol dehydrogenase signature.

CONSENSUS: I-D-x(2)-A-G-P-[ST]-E-[LIVS]-[LIVMA](3)-[AC]-x(3)-A-x(4)-[LIVM]-[AV]-

CONSENSUS: [SACL]-[DE]-[LIVMFC]-[LIVM]-[SA]-x(2)-E-H.

NAME: L-lactate dehydrogenase active site.

CONSENSUS: [LIVMA]-G-[EQ]-H-G-[DN]-[ST].

NAME: D-isomer specific 2-hydroxyacid dehydrogenases NAD-binding signature.

CONSENSUS: [LIVMA]-[AG]-[IVT]-[LIVMFY]-[AG]-x-G-[NHKRQGSAC]-[LIV]-G-x(13,14)-

CONSENSUS: [LIVMT]-x(2)-[FYwCTH]-[DNSTK].

NAME: D-isomer specific 2-hydroxyacid dehydrogenases signature 2.

CONSENSUS: [LIVMFYWA]-[LIVFYWC]-x(2)-[SAC]-[DNQHR]-[IVFA]-[LIVF]-x-[LIVF]-[HNI]-x-

CONSENSUS: P-x(4)-[STN]-x(2)-[LIVMF]-x-[GSDN].

NAME: D-isomer specific 2-hydroxyacid dehydrogenases signature 3.

CONSENSUS: [LMFATC]-[KPQ]-x-[GSTDN]-x-[LIVMFYWR]-[LIVMFYW](2)-N-x-[STAGC]-R-[GP]-x-

CONSENSUS: [LIVH]-[LIVMC]-[DNV].

NAME: 3-hydroxyisobutyrate dehydrogenase signature.

CONSENSUS: [LIVMFY](2)-G-L-G-x-[MQ]-G-x-[PGS]-[MA]-[SA].

NAME: Hydroxymethylglutaryl-coenzyme A reductases signature 1.

CONSENSUS: [RKH]-x(6)-D-x-M-G-x-N-x-[LIVMA].

NAME: Hydroxymethylglutaryl-coenzyme A reductases signature 2.

CONSENSUS: [LIVM]-G-x-[LIVM]-G-G-[AG]-T.

NAME: Hydroxymethylglutaryl-coenzyme A reductases signature 3.

CONSENSUS: A-[LIVM]-x-[STAN]-x(2)-[LI]-x-[KRNQ]-[GSA]-H-[LM]-x-[FYLH].

NAME: Hydroxymethylglutaryl-coenzyme A reductases profile.

NAME: 3-hydroxyacyl-CoA dehydrogenase signature.

CONSENSUS: [DNE]-x(2)-[GA]-F-[LIVMFY]-x-[NT]-R-x(3)-[PA]-[LIVMFY](2)-x(5)-

CONSENSUS: [LIVMFYCT]-[LIVMFY]-x(2)-[GV].

NAME: Malate dehydrogenase active site signature.

CONSENSUS: [LIVM]-T-[TRKMN]-L-D-x(2)-R-[STA]-x(3)-[LIVMFY].

NAME: Malic enzymes signature.

CONSENSUS: F-x-[DV]-D-x(2)-G-T-[GSA]-x-[IV]-x-[LIVMA]-[GAST](2)-[LIVMF](2).

NAME: Isocitrate and isopropylmalate dehydrogenases signature.

CONSENSUS: [NS]-[LIMYT]-[FYDN]-G-[DNT]-[IMVY]-x-[STGDN]-[DN]-x(2)-[SGAP]-x(3,4)-G-

CONSENSUS: [STG]-[LIVMPA]-G-[LIVMF].

NAME: 6-phosphogluconate dehydrogenase signature.

CONSENSUS: [LIVM]-x-D-x(2)-[GA]-[NQS]-K-G-T-G-x-W.

NAME: Glucose-6-phosphate dehydrogenase active site.

CONSENSUS: D-H-Y-L-G-K-[EQK].

NAME: IMP dehydrogenase / GMP reductase signature.

CONSENSUS: [LIVM]-[RK]-[LIVM]-G-[LIVM]-G-x-G-S-[LIVM]-C-x-T.

NAME: Bacterial quinoprotein dehydrogenases signature 1.

CONSENSUS: [DEN]-W-x(3)-G-[RK]-x(6)-[FYW]-S-x(4)-[LIVM]-N-x(2)-N-V-x(2)-L-[RK].

NAME: Bacterial quinoprotein dehydrogenases signature 2.

CONSENSUS: W-x(4)-Y-D-x(3)-[DN]-[LIVMFY](4)-x(2)-G-x(2)-[STA]-P.

NAME: FMN-dependent alpha-hydroxy acid dehydrogenases active site.

CONSENSUS: S-N-H-G-[AG]-R-Q.

NAME: GMC oxidoreductases signature 1.

CONSENSUS: [GA]-[RKN]-x-[LIV]-G(2)-[GST](2)-x-[LIVM]-N-x(3)-[FYWA]-x(2)-[PAG]-x(5)-

CONSENSUS: [DNESH].

NAME: GMC oxidoreductases signature 2.

CONSENSUS: [GS]-[PSTA]-x(2)-[ST]-P-x-[LIVM](2)-x(2)-S-G-[LIVM]-G.

NAME: Eukaryotic molybdopterin oxidoreductases signature.

CONSENSUS: [GA]-x(3)-[KRNQHT]-x(11,14)-[LIVMFYWS]-x(8)-[LIVMF]-x-C-x(2)-[DEN]-R-

CONSENSUS: x(2)-[DE].

NAME: Prokaryotic molybdopterin oxidoreductases signature 1.

CONSENSUS: [STAN]-x-[CH]-x(2,3)-C-[STAG]-[GSTVMF]-x-C-x-[LIVMFYW]-x-[LIVMA]-x(3,4)-

CONSENSUS: [DENQKHT].

NAME: Prokaryotic molybdopterin oxidoreductases signature 2.

CONSENSUS: [STA]-x-[STAC](2)-x(2)-[STA]-D-[LIVMY](2)-L-P-x-[STAC](2)-x(2)-E.

NAME: Prokaryotic molybdopterin oxidoreductases signature 3.

CONSENSUS: A-x(3)-[GDT]-I-x-[DNQTK]-x-[DEA]-x-[LIVM]-x-[LIVMC]-x-[NS]-x(2)-[GS]-

CONSENSUS: x(5)-A-x-[LIVM]-[ST].

NAME: Aldehyde dehydrogenases glutamic acid active site.
 CONSENSUS: [LIVMFGA]-E-[LIMSTAC]-[GS]-G-[KNLM]-[SADN]-[TAPFV].

NAME: Aldehyde dehydrogenases cysteine active site.
 CONSENSUS: [FYLVA]-x(3)-G-[QE]-x-C-[LIVMGSTANC]-[AGCN]-x-[GSTADNEKR].

NAME: Aspartate-semialdehyde dehydrogenase signature.
 CONSENSUS: [LIVM]-[SADN]-x(2)-C-x-R-[LIVM]-x(4)-[GSC]-H-[STA].

NAME: Glyceraldehyde 3-phosphate dehydrogenase active site.
 CONSENSUS: [ASV]-S-C-[NT]-T-x(2)-[LIM].

NAME: N-acetyl-gamma-glutamyl-phosphate reductase active site.
 CONSENSUS: [LIVM]-[GSA]-x-P-G-C-[FY]-[AVP]-T-[GA]-x(3)-[GTAC]-[LIVM]-x-P.

NAME: Gamma-glutamyl phosphate reductase signature.
 CONSENSUS: V-x(5)-A-[LIV]-x-H-I-x(2)-[HY]-[GS]-[ST]-x-H-[ST]-[DE]-x-I.

NAME: Dihydrodipicolinate reductase signature.
 CONSENSUS: E-[IV]-x-E-x-H-x(3)-K-x-D-x-P-S-G-T-A.

NAME: Dihydroorotate dehydrogenase signature 1.
 CONSENSUS: [GS]-x(4)-[GK]-[STA]-[IVSTA]-[GT]-x(3)-[NQR]-x-G-[NH]-x(2)-P-[RT].

NAME: Dihydroorotate dehydrogenase signature 2.
 CONSENSUS: [LIV](2)-[GSA]-x-G-G-[IV]-x-[STGN]-x(3)-[ACV]-x(6)-G-A.

NAME: Coproporphyrinogen III oxidase signature.
 CONSENSUS: K-x-W-C-x(2)-[FYH](3)-[LIVM]-x-H-R-x-E-x-R-G-[LIVM]-G-G-[LIVM]-F-F-D.

NAME: Fumarate reductase / succinate dehydrogenase FAD-binding site.
 CONSENSUS: R-[ST]-H-[ST]-x(2)-A-x-G-G.

NAME: Acyl-CoA dehydrogenases signature 1.
 CONSENSUS: [GAC]-[LIVM]-[ST]-E-x(2)-[GSAN]-G-[ST]-D-x(2)-[GSA].

NAME: Acyl-CoA dehydrogenases signature 2.
 CONSENSUS: [QDE]-x(2)-G-[GS]-x-G-[LIVMFY]-x(2)-[DEN]-x(4)-[KR]-x(3)-[DEN].

NAME: Alanine dehydrogenase & pyridine nucleotide transhydrogenase signature 1.
 CONSENSUS: G-[LIVM]-P-x-E-x(3)-N-E-x(1,3)-R-V-A-x-[ST]-P-x-[GST]-V-x(2)-L-x-[KRH]-x-G.
 CONSENSUS: x-G.

NAME: Alanine dehydrogenase & pyridine nucleotide transhydrogenase signature 2.
 CONSENSUS: [LIVM](2)-G-[GA]-G-x-A-G-x(2)-[SA]-x(3)-[GA]-x-[SG]-[LIVM]-G-A-x-V-x(3)-D.
 CONSENSUS: x(3)-D.

NAME: Glu / Leu / Phe / Val dehydrogenases active site.
 CONSENSUS: [LIV]-x(2)-G-G-[SAG]-K-x-[GV]-x(3)-[DNST]-[PL].

NAME: D-amino acid oxidases signature.
 CONSENSUS: [LIVM](2)-H-[NHA]-Y-G-x-[GSA](2)-x-G-x(5)-G-x-A.

NAME: Pyridoxamine 5'-phosphate oxidase signature.
 CONSENSUS: [LIVF]-E-F-W-[QHG]-x(4)-R-[LIVM]-H-[DNE]-R.

NAME: Copper amine oxidase topaquinone signature.
 CONSENSUS: [LIVM]-[LIVMA]-[LIVM]-x(4)-T-x(2)-N-Y-[DE]-[YN].

NAME: Copper amine oxidase copper-binding site signature.
 CONSENSUS: T-x-G-x(2)-H-[LIVMF]-x(3)-E-[DE]-x-P.

NAME: Lysyl oxidase putative copper-binding region signature.
 CONSENSUS: W-E-W-H-S-C-H-Q-H-Y-H.

NAME: Delta 1-pyrroline-5-carboxylate reductase signature.
 CONSENSUS: [PALF]-x(2,3)-[LIV]-x(3)-[LIVM]-[STAC]-[STV]-x-[GAN]-G-x-T-x(2)-[AG]-[LIV]-x(2)-[LMF]-[DENQK].
 CONSENSUS: [LIV]-x(2)-[LMF]-[DENQK].

NAME: Dihydrofolate reductase signature.
 CONSENSUS: [LVAGC]-[LIF]-G-x(4)-[LIVMF]-P-W-x(4,5)-[DE]-x(3)-[FYTV]-x(3)-[STIQ].

NAME: Tetrahydrofolate dehydrogenase/cyclohydrolase signature 1.
 CONSENSUS: [EQ]-x-[EQK]-[LIVM](2)-x(2)-[LIVM]-x(2)-[LIVMY]-N-x-[DN]-x(5)-[LIVMF](3)-

CONSENSUS: Q-L-P-[LV].

NAME: Tetrahydrofolate dehydrogenase/cyclohydrolase signature 2.

CONSENSUS: P-G-G-V-G-P-[MF]-T-[IV].

NAME: Oxygen oxidoreductases covalent FAD-binding site.

CONSENSUS: P-x(10)-[DE]-[LIVM]-x(3)-[LIVM]-x(9)-[LIVM]-x(3)-[GSA]-[GST]-G-H.

NAME: Pyridine nucleotide-disulphide oxidoreductases class-I active site.

CONSENSUS: G-G-x-C-[LIVA]-x(2)-G-C-[LIVM]-P.

NAME: Pyridine nucleotide-disulphide oxidoreductases class-II active site.

CONSENSUS: C-x(2)-C-D-[GA]-x(2,4)-[FY]-x(4)-[LIVM]-x-[LIVM](2)-G(3)-[DN].

NAME: Respiratory-chain NADH dehydrogenase subunit 1 signature 1.

CONSENSUS: G-[LIVMFYKRS]-[LIVMAGP]-Q-x-[LIVMFY]-x-D-[AGIM]-[LIVMFTA]-K-[LVMYST]-

CONSENSUS: [LIVMFYG]-x-[KR]-[EQG].

NAME: Respiratory-chain NADH dehydrogenase subunit 1 signature 2.

CONSENSUS: P-F-D-[LIVMFYQ]-[STAGPVM]-E-[GAC]-E-x-[EQ]-[LIVMS]-x(2)-G.

NAME: Respiratory-chain NADH dehydrogenase 20 Kd subunit signature.

CONSENSUS: [GN]-x-D-[KRST]-[LIVMF](2)-P-[IV]-D-[LIVMFYW](2)-x-P-x-C-P-[PT].

NAME: Respiratory-chain NADH dehydrogenase 24 Kd subunit signature.

CONSENSUS: D-x(2)-F-[ST]-x(5)-C-L-G-x-C-x(2)-[GA]-P.

NAME: Respiratory chain NADH dehydrogenase 30 Kd subunit signature.

CONSENSUS: E-R-E-x(2)-[DE]-[LIVMF](2)-x(6)-[HK]-x(3)-[KRP]-x-[LIVM]-[LIVMS].

NAME: Respiratory chain NADH dehydrogenase 49 Kd subunit signature.

CONSENSUS: [LIVMH]-H-[RT]-[GA]-x-E-K-[LIVMT]-x-E-x-[KRQ].

NAME: Respiratory-chain NADH dehydrogenase 51 Kd subunit signature 1.

CONSENSUS: G-[AM]-G-[AR]-Y-[LIVM]-C-G-[DE](2)-[STA](2)-[LIM](2)-[EN]-S.

NAME: Respiratory-chain NADH dehydrogenase 51 Kd subunit signature 2.

CONSENSUS: E-S-C-G-x-C-x-P-C-R-x-G.

NAME: Respiratory-chain NADH dehydrogenase 75 Kd subunit signature 1.

CONSENSUS: P-x(2)-C-[YWS]-x(7)-G-x-C-R-x-C.

NAME: Respiratory-chain NADH dehydrogenase 75 Kd subunit signature 2.

CONSENSUS: C-P-x-C-[DE]-x-[GS](2)-x-C-x-L-Q.

NAME: Respiratory-chain NADH dehydrogenase 75 Kd subunit signature 3.

CONSENSUS: R-C-[LIVM]-x-C-x-R-C-[LIVM]-x-[FY].

NAME: Nitrite and sulfite reductases iron-sulfur/siroheme-binding site.

CONSENSUS: [STV]-G-C-x(3)-C-x(6)-[DE]-[LIVMF]-[GAT]-[LIVMF].

NAME: Uricase signature.

CONSENSUS: L-x-[LV]-L-K-[ST]-T-x-S-x-F-x(2)-[FY]-x(4)-[FY].

NAME: Heme-copper oxidase catalytic subunit, copper B binding region signature.

CONSENSUS: [YWG]-[LIVFYWTA](2)-[VGS]-H-[LNP]-x-V-x(44,47)-H-H.

NAME: CO II and nitrous oxide reductase dinuclear copper centers signature.

CONSENSUS: V-x-H-x(33,40)-C-x(3)-C-x(3)-H-x(2)-M.

NAME: Cytochrome c oxidase subunit Vb, zinc binding region signature.

CONSENSUS: [LIVM](2)-[FYW]-x(10)-C-x(2)-C-G-x(2)-[FY]-K-L.

NAME: Multicopper oxidases signature 1.

CONSENSUS: G-x-[FYW]-x-[LIVMFYW]-x-[CST]-x(8)-G-[LM]-x(3)-[LIVMFYW].

NAME: Multicopper oxidases signature 2.

CONSENSUS: H-C-H-x(3)-H-x(3)-[AG]-[LM].

NAME: Peroxidases proximal heme-ligand signature.

CONSENSUS: [DET]-[LIVMTA]-x(2)-[LIVM]-[LIVMSTAG]-[SAG]-[LIVMSTAG]-H-[STA]-[LIVMFY].

NAME: Peroxidases active site signature.

CONSENSUS: [SGATV]-x(3)-[LIVMA]-R-[LIVMA]-x-[FW]-H-x-[SAC].

NAME: Catalase proximal heme-ligand signature.
 CONSENSUS: R-[LIVMFSTAN]-F-[GASTNP]-Y-x-D-[AST]-[QEH].

NAME: Catalase proximal active site signature.
 CONSENSUS: [IF]-x-[RH]-x(4)-[EQ]-R-x(2)-H-x(2)-[GAS]-[GASTF]-[GAST].

NAME: Glutathione peroxidases selenocysteine active site.
 CONSENSUS: [GN]-[RKHNFCY]-x-[LIVMFC]-[LIVMF](2)-x-N-[VT]-x-[STC]-x-C-[GA]-x-T.

NAME: Glutathione peroxidases signature 2.
 CONSENSUS: [LIV]-[AGD]-F-P-[CS]-[NG]-Q-F.

NAME: Lipoygenases iron-binding region signature 1.
 CONSENSUS: H-[EQ]-x(3)-H-x-[LM]-[NQRC]-[GST]-H-[LIVMSTAC](3)-E.

NAME: Lipoygenases iron-binding region signature 2.
 CONSENSUS: [LIVMA]-H-P-[LIVM]-x-[KRQ]-[LIVMF](2)-x-[AP]-H.

NAME: Extradiol ring-cleavage dioxygenases signature.
 CONSENSUS: [GNTIV]-x-H-x(5,7)-[LIVMF]-Y-x(2)-[DENTA]-P-x-[GP]-x(2,3)-E.

NAME: Intradiol ring-cleavage dioxygenases signature.
 CONSENSUS: [LIVM]-x-G-x-[LIVM]-x(4)-[GS]-x(2)-[LIVM]-x(4)-[LIVM]-[DE]-[LIVMFY]-x(6)-G-x-[FY].

NAME: Indoleamine 2,3-dioxygenase signature 1.
 CONSENSUS: G-G-S-[AN]-[GA]-Q-S-S-x(2)-Q.

NAME: Indoleamine 2,3-dioxygenase signature 2.
 CONSENSUS: [FY]-L-[DQ]-[DE]-[LIVM]-x(2)-Y-M-x(3)-H-[KR].

NAME: Bacterial ring hydroxylating dioxygenases alpha-subunit signature.
 CONSENSUS: C-x-H-R-[GA]-x(8)-G-N-x(5)-C-x-[FY]-H.

NAME: Bacterial luciferase subunits signature.
 CONSENSUS: [GA]-[LIVM]-P-[LIVM]-x-[LIVMFY]-x-W-x(6)-[RK]-x(6)-Y-x(3)-[AR].

NAME: ubiH/COQ6 monooxygenase family signature.
 CONSENSUS: H-P-[LIV]-[AG]-G-Q-G-x-N-x-G-x(2)-D.

NAME: Biotpterin-dependent aromatic amino acid hydroxylases signature.
 CONSENSUS: P-D-x(2)-H-[DE]-[LI]-[LIVMF]-G-H-[LIVMC]-P.

NAME: Copper type II, ascorbate-dependent monooxygenases signature 1.
 CONSENSUS: H-H-M-x(2)-F-x-C.

NAME: Copper type II, ascorbate-dependent monooxygenases signature 2.
 CONSENSUS: H-x-F-x(4)-H-T-H-x(2)-G.

NAME: Tyrosinase CuA-binding region signature.
 CONSENSUS: H-x(4,5)-F-[LIVMFTP]-x-[FW]-H-R-x(2)-[LM]-x(3)-E.

NAME: Tyrosinase and hemocyanins CuB-binding region signature.
 CONSENSUS: D-P-x-F-[LIVMFYW]-x(2)-H-x(3)-D.

NAME: Fatty acid desaturases family 1 signature.
 CONSENSUS: G-E-x-[FY]-H-N-[FY]-H-H-x-F-P-x-D-Y.

NAME: Fatty acid desaturases family 2 signature.
 CONSENSUS: [ST]-[SA]-x(3)-[QR]-[LI]-x(5,6)-D-Y-x(2)-[LIVMFYW]-[LIVM]-[DE].

NAME: Cytochrome P450 cysteine heme-iron ligand signature.
 CONSENSUS: [FW]-[SGNH]-x-[GD]-x-[RHPT]-x-C-[LIVMFAP]-[GAD].

NAME: Heme oxygenase signature.
 CONSENSUS: L-L-V-A-H-A-Y-T-R.

NAME: Copper/Zinc superoxide dismutase signature 1.
 CONSENSUS: [GA]-[IFAT]-H-[LIVF]-H-x(2)-[GP]-[SDG]-x-[STAGD].

NAME: Copper/Zinc superoxide dismutase signature 2.
 CONSENSUS: G-[GN]-[SGA]-G-x-R-x-[SGA]-C-x(2)-[IV].

NAME: Manganese and iron superoxide dismutases signature.

CONSENSUS: D-x-W-E-H-[STA]-[FY](2).

NAME: Ribonucleotide reductase large subunit signature.

CONSENSUS: W-x(2)-[LF]-x(6,7)-G-[LIVM]-[FYRA]-[NH]-x(3)-[STAQLIVM]-[ASC]-x(2)-

CONSENSUS: [PA].

NAME: Ribonucleotide reductase small subunit signature.

CONSENSUS: [IVMSEQ]-E-x(1,2)-[LIVTA]-[HY]-[GSA]-x-[STAVM]-Y-x(2)-[LIVMQ]-x(3)-

CONSENSUS: [LIFY]-[IVFYCSA].

NAME: Nitrogenases component 1 alpha and beta subunits signature 1.

CONSENSUS: [LIVMFYH]-[LIVMFST]-H-[AG]-[AGSP]-[LIVMNQA]-[AG]-C.

NAME: Nitrogenases component 1 alpha and beta subunits signature 2.

CONSENSUS: [STANQ]-[ET]-C-x(5)-G-D-[DN]-[LIVMT]-x-[STAGR]-[LIVMFYST].

NAME: NifH/frxC family signature 1.

CONSENSUS: E-x-G-G-P-x(2)-[GA]-x-G-C-[AG]-G.

NAME: NifH/frxC family signature 2.

CONSENSUS: D-x-L-G-D-V-V-C-G-G-F-[AG]-x-P.

NAME: Nickel-dependent hydrogenases large subunit signature 1.

CONSENSUS: R-G-[LIVMF]-E-x(15)-[QESM]-R-x-C-G-[LIVM]-C.

NAME: Nickel-dependent hydrogenases large subunit signature 2.

CONSENSUS: [FY]-D-P-C-[LIM]-[ASG]-C-x(2,3)-H.

NAME: Glutamyl-tRNA reductase signature.

CONSENSUS: H-[LIVM]-x(2)-[LIVM]-[GSTAC](3)-[LIVM]-[DEQ]-S-[LIVMA]-[LIVM](2)-[GF]-E-

CONSENSUS: x-[QR]-[IV]-[LIT]-[STAG]-Q-[LIVM]-[KR].

NAME: Bacterial-type phytoene dehydrogenase signature.

CONSENSUS: [NG]-x-[FYWV]-[LIVMF]-x-G-[AGC]-[GS]-[TA]-[HQT]-P-G-[STAV]-G-[LIVM]-

CONSENSUS: x(5)-[GS].

NAME: Glycine radical signature.

CONSENSUS: [STIV]-x-R-[IVT]-[CSA]-G-Y-x-[GACV].

NAME: Ergosterol biosynthesis ERG4/ERG24 family signature 1.

CONSENSUS: G-x(2)-[LIVM]-Y-D-x-[FY]-x-G-x(2)-L-N-P-R.

NAME: Ergosterol biosynthesis ERG4/ERG24 family signature 2.

CONSENSUS: [LIVM](2)-H-R-x(2)-R-D-x(3)-C-x(2)-K-Y-G.

NAME: NNMT/PNMT/TEMT family of methyltransferases signature.

CONSENSUS: L-I-D-I-G-S-G-P-T-[IV]-Y-Q-L-L-S-A-C.

NAME: RNA methyltransferase trmA family signature 1.

CONSENSUS: [DN]-P-[PA]-R-x-G-x(14,16)-[LIVM](2)-Y-x-S-C-N-x(2)-T.

NAME: RNA methyltransferase trmA family signature 2.

CONSENSUS: [LIVMF]-D-x-F-P-[QHY]-[ST]-x-H-[LIVMFY]-E.

NAME: Thymidylate synthase active site.

CONSENSUS: R-x(2)-[LIVM]-x(3)-[FW]-[QN]-x(8,9)-[LV]-x-P-C-[HAVM]-x(3)-[QMT]-[FYW]-

CONSENSUS: x-[LV].

NAME: Ribosomal RNA adenine dimethylases signature.

CONSENSUS: [LIVM]-[LIVMFY]-[DE]-x-G-[STAPV]-G-x-[GA]-x-[LIVMF]-[ST]-x(2)-[LIVM]-

CONSENSUS: x(6)-[LIVMY]-x-[STAGV]-[LIVMFYHC]-E-x-D.

NAME: Methylated-DNA-protein-cysteine methyltransferase active site.

CONSENSUS: [LIVMF]-P-C-H-R-[LIVMF](2).

NAME: N-6 Adenine-specific DNA methylases signature.

CONSENSUS: [LIVMAC]-[LIVFYWA]-x-[DN]-P-P-[FYW].

NAME: N-4 cytosine-specific DNA methylases signature.

CONSENSUS: [LIVMF]-T-S-P-P-[FY].

NAME: C-5 cytosine-specific DNA methylases active site.

CONSENSUS: [DENKS]-x-[FLIV]-x(2)-[GSTC]-x-P-C-x(2)-[FYWLIM]-S.

NAME: C-5 cytosine-specific DNA methylases C-terminal signature.
 CONSENSUS: [RKQGT]-x(2)-G-N-[STAG]-[LIVMF]-x(3)-[LIVMT]-x(3)-[LIVM]-x(3)-[LIVM].

NAME: Protein-L-isoaspartate(D-aspartate) O-methyltransferase signature.
 CONSENSUS: [GSA]-D-G-x(2)-G-[FYWV]-x(3)-[AS]-P-[FY]-[DN]-x-I.

NAME: Uroporphyrin-III C-methyltransferase signature 1.
 CONSENSUS: [LIVM]-[GS]-[STAL]-G-P-G-x(3)-[LIVMFY]-[LIVM]-T-[LIVM]-[KRHQG]-[AG].

NAME: Uroporphyrin-III C-methyltransferase signature 2.
 CONSENSUS: V-x(2)-[LI]-x(2)-G-D-x(3)-[FYW]-[GS]-x(8)-[LIVF]-x(5,6)-[LIVMFYWPAC]-
 CONSENSUS: x-[LIVMY]-x-P-G.

NAME: ubiE/COQ5 methyltransferase family signature 1.
 CONSENSUS: Y-D-x-M-N-x(2)-[LIVM]-S-x(3)-H-x(2)-W.

NAME: ubiE/COQ5 methyltransferase family signature 2.
 CONSENSUS: R-V-[LIVM]-K-[PV]-G-G-x-[LIVMF]-x(2)-[LIVM]-E-x-S.

NAME: Serine hydroxymethyltransferase pyridoxal-phosphate attachment site.
 CONSENSUS: [DEH]-[LIVMFY]-x-[STMV]-[GST]-[ST](2)-H-K-[ST]-[LF]-x-G-[PAC]-[RQ]-
 CONSENSUS: [GSA]-[GA].

NAME: Phosphoribosylglycinamide formyltransferase active site.
 CONSENSUS: G-x-[STM]-[IVT]-x-[FYWVQ]-[VMAT]-x-[DEVM]-x-[LIVMY]-D-x-G-x(2)-[LIVT]-
 CONSENSUS: x(6)-[LIVM].

NAME: Aspartate and ornithine carbamoyltransferases signature.
 CONSENSUS: F-x-[EK]-x-S-[GT]-R-T.

NAME: Transketolase signature 1.
 CONSENSUS: R-x(3)-[LIVMTA]-[DENQSTHKF]-x(5,6)-[GSN]-G-H-[PLIVMF]-[GSTA]-x(2)-
 CONSENSUS: [LMC]-[GS].

NAME: Transketolase signature 2.
 CONSENSUS: G-[DEQGS]-[DN]-G-[PAEQ]-[ST]-[HQ]-x-[PAGM]-[LIVMYAC]-[DEFYW]-x(2)-
 CONSENSUS: [STAP]-x(2)-[RGA].

NAME: Transaldolase signature 1.
 CONSENSUS: [DG]-[IVSA]-T-[ST]-N-P-[STA]-[LIVMF](2).

NAME: Transaldolase active site.
 CONSENSUS: [LIVM]-x-[LIVM]-K-[LIVM]-[PAS]-x-[ST]-x-[DENQPAS]-G-[LIVM]-x-[AGV]-x-
 CONSENSUS: [QEKIRST]-x-[LIVM].

NAME: Acyltransferases ChoActase / COT / CPT family signature 1.
 CONSENSUS: [LI]-P-x-[LVP]-P-[IVTA]-P-x-[LIVM]-x-[DENQAS]-[ST]-[LIVM]-x(2)-[LY].

NAME: Acyltransferases ChoActase / COT / CPT family signature 2.
 CONSENSUS: R-[FYW]-x-[DA]-[KA]-x(0,1)-[LIVMFY]-x-[LIVMFY](2)-x(3)-[DNS]-[GSA]-x(6)-
 CONSENSUS: [DE]-[HS]-x(3)-[DE]-[GA].

NAME: Thiolases acyl-enzyme intermediate signature.
 CONSENSUS: [LIVM]-[NST]-x(2)-C-[SAGL]-[ST]-[SAG]-[LIVMFYNS]-x-[STAG]-[LIVM]-x(6)-
 CONSENSUS: [LIVM].

NAME: Thiolases signature 2.
 CONSENSUS: N-x(2)-G-G-x-[LIVM]-[SA]-x-G-H-P-x-G-x-[ST]-G.

NAME: Thiolases active site.
 CONSENSUS: [AG]-[LIVMA]-[STAGLIVM]-[STAG]-[LIVMA]-C-x-[AG]-x-[AG]-x-[AG]-x-[SAG].

NAME: Chloramphenicol acetyltransferase active site.
 CONSENSUS: Q-[LIV]-H-H-[SA]-x(2)-D-G-[FY]-H.

NAME: Hexapeptide-repeat containing-transferases signature.
 CONSENSUS: [LIV]-[GAED]-x(2)-[STAV]-x-[LIV]-x(3)-[LIVAC]-x-[LIV]-[GAED]-x(2)-
 CONSENSUS: [STAVR]-x-[LIV]-[GAED]-x(2)-[STAV]-x-[LIV]-x(3)-[LIV].

NAME: Beta-ketoacyl synthases active site.
 CONSENSUS: G-x(4)-[LIVMFAP]-x(2)-[AGC]-C-[STA](2)-[STAG]-x(3)-[LIVMF].

NAME: Chalcone and stilbene synthases active site.

CONSENSUS: R-[LIVMFYS]-x-[LIVM]-x-[QHG]-x-G-C-[FYNA]-[GA]-G-[GA]-[STAV]-x-[LIVMF]-
 CONSENSUS: [RA].

NAME: Myristoyl-CoA:protein N-myristoyltransferase signature 1.
 CONSENSUS: E-I-N-F-L-C-x-H-K.

NAME: Myristoyl-CoA:protein N-myristoyltransferase signature 2.
 CONSENSUS: K-F-G-x-G-D-G.

NAME: Gamma-glutamyltranspeptidase signature.
 CONSENSUS: T-[STA]-H-x-[ST]-[LIVMA]-x(4)-G-[SN]-x-V-[STA]-x-T-x-T-[LIVM]-[NE]-
 CONSENSUS: x(1,2)-[FY]-G.

NAME: Transglutaminases active site.
 CONSENSUS: [GT]-Q-[CA]-W-V-x-[SA]-[GA]-[IVT]-x(2)-T-x-[LMSC]-R-[CSA]-[LV]-G.

NAME: Phosphorylase pyridoxal-phosphate attachment site.
 CONSENSUS: E-A-[SC]-G-x-[GS]-x-M-K-x(2)-[LM]-N.

NAME: UDP-glycosyltransferases signature.
 CONSENSUS: [FW]-x(2)-Q-x(2)-[LIVMYA]-[LIMV]-x(4,6)-[LVGAC]-[LVFYA]-[LIVMF]-[STAGCM]-
 CONSENSUS: [HNQ]-[STAGC]-G-x(2)-[STAG]-x(3)-[STAGL]-[LIVMFA]-x(4)-[PQR]-[LIVMT]-
 CONSENSUS: x(3)-[PA]-x(3)-[DES]-[QEHN].

NAME: Purine/pyrimidine phosphoribosyl transferases signature.
 CONSENSUS: [LIVMFYWCTA]-[LIVM]-[LIVMA]-[LIVMFC]-[DE]-D-[LIVMS]-[LIVM]-[STAVD]-
 CONSENSUS: [STAR]-[GAC]-x-[STAR].

NAME: Glutamine amidotransferases class-I active site.
 CONSENSUS: [PAS]-[LIVMFYT]-[LIVMFY]-G-[LIVMFY]-C-[LIVMFYN]-G-x-[QEH]-x-[LIVMFA].

NAME: Glutamine amidotransferases class-II active site.
 CONSENSUS: < x(0,11)-C-[GS]-[IV]-[LIVMFYW]-[AG].

NAME: Purine and other phosphorylases family 1 signature.
 CONSENSUS: [GST]-x-G-[LIVM]-G-x-[PA]-S-x-[GSTA]-I-x(3)-E-L.

NAME: Purine and other phosphorylases family 2 signature.
 CONSENSUS: [LIV]-x(3)-G-x(2)-H-x-[LIVMFY]-x(4)-[LIVMF]-x(3)-[ATV]-x(1,2)-[LIVM]-x-
 CONSENSUS: [ATV]-x(4)-[GN]-x(3,4)-[LIVMF](2)-x(2)-[STN]-[SA]-x-G-[GS]-[LIVM].

NAME: Thymidine and pyrimidine-nucleoside phosphorylases signature.
 CONSENSUS: S-[GS]-R-[GA]-[LIV]-x(2)-[TA]-[GA]-G-T-x-D-x-[LIV]-E.

NAME: ATP phosphoribosyltransferase signature.
 CONSENSUS: E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-G-x-T-[LM].

NAME: NAD:arginine ADP-ribosyltransferases signature.
 CONSENSUS: [FY]-x-[FY]-K-x(2)-H-[FY]-x-L-[ST]-x-A.

NAME: Prolipoprotein diacylglycerol transferase signature.
 CONSENSUS: G-R-x-[GA]-N-F-[LIVMF]-N-x-E-x(2)-G.

NAME: S-adenosylmethionine synthetase signature 1.
 CONSENSUS: G-A-G-D-Q-G-x(3)-G-Y.

NAME: S-adenosylmethionine synthetase signature 2.
 CONSENSUS: G-[GA]-G-[ASC]-F-S-x-K-[DE].

NAME: Polyprenyl synthetases signature 1.
 CONSENSUS: [LIVM](2)-x-D-D-x(2,4)-D-x(4)-R-R-[GH].

NAME: Polyprenyl synthetases signature 2.
 CONSENSUS: [LIVMFY]-G-x(2)-[FY]-Q-[LIVM]-x-D-D-[LIVMFY]-x-[DNG].

NAME: Squalene and phytoene synthases signature 1.
 CONSENSUS: Y-[CSAM]-x(2)-[VSG]-A-[GSA]-[LIVAT]-[IV]-G-x(2)-[LMSC]-x(2)-[LIV].

NAME: Squalene and phytoene synthases signature 2.
 CONSENSUS: [LIVM]-G-x(3)-Q-x(2,3)-N-[IF]-x-R-D-[LIVMFY]-x(2)-[DE]-x(4,7)-R-x-[FY]-
 CONSENSUS: x-P.

NAME: Protein prenyltransferases alpha subunit repeat signature.
 CONSENSUS: [PSIAV]-x-[NDFV]-[NEQIY]-x-[LIVMAGP]-W-[NQSTHF]-[FYHQ]-[LIVMR].

NAME: Riboflavin synthase alpha chain family signature.
 CONSENSUS: [LIVMF]-x(5)-G-[STADNQ]-[KREQIYW]-V-N-[LIVM]-E.

NAME: Dihydropteroate synthase signature 1.
 CONSENSUS: [LIVM]-x-[AG]-[LIVMF](2)-N-x-T-x-D-S-F-x-D-x-[SG].

NAME: Dihydropteroate synthase signature 2.
 CONSENSUS: [GE]-[SA]-x-[LIVM](2)-D-[LIVM]-G-[GP]-x(2)-[STA]-x-P.

NAME: EPSP synthase signature 1.
 CONSENSUS: [LIVM]-x(2)-[GN]-N-[SA]-G-T-[STA]-x-R-x-[LIVMY]-x-[GSTA].

NAME: EPSP synthase signature 2.
 CONSENSUS: [KR]-x-[KH]-E-[CST]-[DNE]-R-[LIVM]-x-[STA]-[LIVMC]-x(2)-[EN]-[LIVMF]-x-
 CONSENSUS: [KRA]-[LIVMF]-G.

NAME: FLAP/GST2/LTC4S family signature.
 CONSENSUS: G-x(3)-F-E-R-V-[FY]-x-A-[NQ]-x-N-C.

NAME: Aminotransferases class-I pyridoxal-phosphate attachment site.
 CONSENSUS: [GS]-[LIVMFYTAC]-[GSTA]-K-x(2)-[GSALVN]-[LIVMFA]-x-[GNAR]-x-R-[LIVMA]-
 CONSENSUS: [GA].

NAME: Aminotransferases class-II pyridoxal-phosphate attachment site.
 CONSENSUS: T-[LIVMFYW]-[STAG]-K-[SAG]-[LIVMFYWR]-[SAG]-x(2)-[SAG].

NAME: Aminotransferases class-III pyridoxal-phosphate attachment site.
 CONSENSUS: [LIVMFYWC](2)-x-D-E-[LIVMA]-x(2)-[GP]-x(0,1)-[LIVMFYWAG]-x(0,1)-[SACR]-x-
 CONSENSUS: [GSAD]-x(12,16)-D-[LIVMFYWC]-x(2,3)-[GSA]-K-x(3)-[GSTADN]-[GSA].

NAME: Aminotransferases class-IV signature.
 CONSENSUS: E-x-[STAGCI]-x(2)-N-[LIVMFAC]-[FY]-x(6,12)-[LIVMF]-x-T-x(6,8)-[LIVM]-x-
 CONSENSUS: [GS]-[LIVM]-x-[KR].

NAME: Aminotransferases class-V pyridoxal-phosphate attachment site.
 CONSENSUS: [LIVFYCHT]-[DGH]-[LIVMFYAC]-[LIVMFYA]-x(2)-[GSTAC]-[GSTA]-[HQR]-K-
 CONSENSUS: x(4,6)-G-x-[GSAT]-x-[LIVMFYSAC].

NAME: Hexokinases signature.
 CONSENSUS: [LIVM]-G-F-[TN]-F-S-[FY]-P-x(5)-[LIVM]-[DNST]-x(3)-[LIVM]-x(2)-W-T-K-x-
 CONSENSUS: [LF].

NAME: Galactokinase signature.
 CONSENSUS: G-R-x-N-[LIV]-I-G-E-H-x-D-Y.

NAME: GHMP kinases putative ATP-binding domain.
 CONSENSUS: [LIVM]-[PK]-x-[GSTA]-x(0,1)-G-L-[GS]-S-S-[GSA]-[GSTAC].

NAME: Phosphofructokinase signature.
 CONSENSUS: [RK]-x(4)-G-H-x-Q-[QR]-G-G-x(5)-D-R.

NAME: pfkB family of carbohydrate kinases signature 1.
 CONSENSUS: [AG]-G-x(0,1)-[GAP]-x-N-x-[STA]-x(6)-[GS]-x(9)-G.

NAME: pfkB family of carbohydrate kinases signature 2.
 CONSENSUS: [DNSK]-[PSTV]-x-[SAG](2)-[GD]-D-x(3)-[SAGV]-[AG]-[LIVMFY]-[LIVMSTAP].

NAME: ROK family signature.
 CONSENSUS: [LIVM]-x(2)-G-[LIVMFCT]-G-x-[GA]-[LIVMFA]-x(8)-G-x(3,5)-[GATP]-x(2)-
 CONSENSUS: G-[RKH].

NAME: Phosphoribulokinase signature.
 CONSENSUS: K-[LIVM]-x-R-D-x(3)-R-G-x-[ST]-x-E.

NAME: Thymidine kinase cellular-type signature.
 CONSENSUS: [GA]-x(1,2)-[DE]-x-Y-x-[STAP]-x-C-[NKR]-x-[CH]-[LIVMFYWH].

NAME: FGGY family of carbohydrate kinases signature 1.
 CONSENSUS: [MFYGS]-x-[PST]-x(2)-K-[LIVMFYW]-x-W-[LIVMF]-x-[DENQTKR]-[ENQH].

NAME: FGGY family of carbohydrate kinases signature 2.
 CONSENSUS: [GSA]-x-[LIVMFYW]-x-G-[LIVM]-x(7,8)-[HDENQ]-[LIVMF]-x(2)-[AS]-[STAIVM]-
 CONSENSUS: [LIVMFY]-[DEQ].

NAME: Protein kinases ATP-binding region signature.

CONSENSUS: [LIV]-G-{P}-G-{P}-[FYWMGSTNH]-[SGA]-[PW]-[LIVCAT]-[PD]-x-[GSTACLIVMFY]-
CONSENSUS: x(5,18)-[LIVMFYWCSTAR]-[AIVP]-[LIVMFAGCKR]-K.

NAME: Serine/Threonine protein kinases active-site signature.

CONSENSUS: [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-K-x(2)-N-[LIVMFYCT](3).

NAME: Tyrosine protein kinases specific active-site signature.

CONSENSUS: [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-[RSTAC]-x(2)-N-[LIVMFYC](3).

NAME: Protein kinase domain profile.

NAME: Casein kinase II regulatory subunit signature.

CONSENSUS: C-P-x-[LIVMY]-x-C-x(5)-L-P-[LIVMC]-G-x(9)-V-[KR]-x(2)-C-P-x-C.

NAME: Pyruvate kinase active site signature.

CONSENSUS: [LIVAC]-x-[LIVM](2)-[SAPCV]-K-[LIV]-E-[NKRST]-x-[DEQH]-[GSTA]-[LIVM].

NAME: Shikimate kinase signature.

CONSENSUS: [KR]-x(2)-E-x(3)-[LIVMF]-x(8,12)-[LIVMF](2)-[SA]-x-G(3)-x-[LIVMF].

NAME: Prokaryotic diacylglycerol kinase signature.

CONSENSUS: E-x-[LIVM]-N-[ST]-[SA]-[LIV]-E-x(2)-V-D.

NAME: Phosphatidylinositol 3- and 4-kinases signature 1.

CONSENSUS: [LIVMFAC]-K-x(1,3)-[DEA]-[DE]-[LIVMC]-R-Q-[DE]-x(4)-Q.

NAME: Phosphatidylinositol 3- and 4-kinases signature 2.

CONSENSUS: [GS]-x-[AV]-x(3)-[LIVM]-x(2)-[FYH]-[LIVM](2)-x-[LIVMF]-x-D-R-H-x(2)-N.

NAME: Acetate and butyrate kinases family signature 1.

CONSENSUS: [LIVM](2)-x-[LIVM]-N-x-G-S-[ST]-S-x-[KE].

NAME: Acetate and butyrate kinases family signature 2.

CONSENSUS: [LIVMA](2)-x(2)-H-x-G-x-G-x-[ST]-[LIVM]-x-[AV]-x(3)-G.

NAME: Phosphoglycerate kinase signature.

CONSENSUS: [KRHGTCV]-[VT]-[LIVMF]-[LIVMC]-R-x-D-x-N-[SACV]-P.

NAME: Aspartokinase signature.

CONSENSUS: [LIVM]-x-K-[FY]-G-G-[ST]-[SC]-[LIVM].

NAME: Glutamate 5-kinase signature.

CONSENSUS: [GSTN]-x(2)-G-x-G-[GC]-[IM]-x-[STA]-K-[LIVM]-x-[SA]-[TCA]-x(2)-[GALV]-
CONSENSUS: x(3)-G.

NAME: ATP:guanido phosphotransferases active site.

CONSENSUS: C-P-x(0,1)-[ST]-N-[IL]-G-T.

NAME: PTS HPR component histidine phosphorylation site signature.

CONSENSUS: G-[LIVM]-H-[STA]-R-[PA]-[GSTA]-[STAM].

NAME: PTS HPR component serine phosphorylation site signature.

CONSENSUS: [GSADE]-[KREQTV]-x(4)-[KRN]-S-[LIVMF](2)-x-[LIVM]-x(2)-[LIVM]-[GAD].

NAME: PTS EIIA domains phosphorylation site signature 1.

CONSENSUS: G-x(2)-[LIVMF](3)-H-[LIVMF]-G-[LIVMF]-x-T-[ALV].

NAME: PTS EIIA domains phosphorylation site signature 2.

CONSENSUS: [DENQ]-x(6)-[LIVMF]-[GA]-x(2)-[LIVM]-A-[LIVM]-P-H-[GAC].

NAME: PTS EIIB domains cysteine phosphorylation site signature.

CONSENSUS: N-[LIVMFY]-x(5)-C-x-T-R-[LIVMF]-x-[LIVMF]-x-[LIVM]-x-[DQ].

NAME: Adenylate kinase signature.

CONSENSUS: [LIVMFYW](3)-D-G-[FYI]-P-R-x(3)-[NQ].

NAME: Nucleoside diphosphate kinases active site.

CONSENSUS: N-x(2)-H-[GA]-S-D-[SA]-[LIVMPKNE].

NAME: Guanylate kinase signature.

CONSENSUS: T-[ST]-R-x(2)-[KR]-x(2)-[DE]-x(2)-G-x(2)-Y-x-[FY]-[LIVMK].

NAME: Guanylate kinase domain profile.

NAME: Phosphoribosyl pyrophosphate synthetase signature.

CONSENSUS: D-[LI]-H-[SA]-x-Q-[IMST]-[QM]-G-[FY]-F-x(2)-P-[LIVMFC]-D.

NAME: 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase signature.

CONSENSUS: G-[PE]-R-x(2)-D-L-D-[LIVM](2).

NAME: Bacteriophage-type RNA polymerase family active site signature 1.

CONSENSUS: P-[LIVM]-x(2)-D-[GA]-[ST]-[AC]-[SN]-[GA]-[LIVMFY]-Q.

NAME: Bacteriophage-type RNA polymerase family active site signature 2.

CONSENSUS: [LIVMF]-x-R-x(3)-K-x(2)-[LIVMF]-M-[PT]-x(2)-Y.

NAME: Eukaryotic RNA polymerase II heptapeptide repeat.

CONSENSUS: Y-[ST]-P-[ST]-S-P-[STANK].

NAME: RNA polymerases beta chain signature.

CONSENSUS: G-x-K-[LIVMFA]-[STAC]-[GSTN]-x-[HSTA]-[GS]-[QNH]-K-G-[IVT].

NAME: RNA polymerases M / 15 Kd subunits signature.

CONSENSUS: F-C-x-[DEKST]-C-[GNK]-[DNSA]-[LIVMH]-[LIVM]-x(8,14)-C-x(2)-C.

NAME: RNA polymerases D / 30 to 40 Kd subunits signature.

CONSENSUS: N-[SGA]-[LIVMF]-R-R-x(9)-[SA]-x(3)-V-x(4)-N-x-[STA]-x(3)-[DN]-E-x-[LI]-

CONSENSUS: [GA]-x-R-[LI]-[GA]-[LIVM](2)-P.

NAME: RNA polymerases H / 23 Kd subunits signature.

CONSENSUS: H-[NEI]-[LIVM]-V-P-x-H-x(2)-[LIVM]-x(2)-[DE].

NAME: RNA polymerases K / 14 to 18 Kd subunits signature.

CONSENSUS: [ST]-x-[FY]-E-x-[AT]-R-x-[LIVM]-[GSA]-x-R-[SA]-x-Q.

NAME: RNA polymerases L / 13 to 16 Kd subunits signature.

CONSENSUS: [DE](2)-H-[ST]-[LIVM]-[GAP]-N-x(11)-V-x-[FM]-x(2)-Y-x(3)-H-P.

NAME: RNA polymerases N / 8 Kd subunits signature.

CONSENSUS: [LIVMF](2)-P-[LIVM]-x-C-F-[ST]-C-G.

NAME: DNA polymerase family A signature.

CONSENSUS: R-x(2)-[GSAV]-K-x(3)-[LIVMFY]-[AGQ]-x(2)-Y-x(2)-[GS]-x(3)-[LIVMA].

NAME: DNA polymerase family B signature.

CONSENSUS: [YA]-[GLIVMSTAC]-D-T-D-[SG]-[LIVMFTC]-x-[LIVMSTAC].

NAME: DNA polymerase family X signature.

CONSENSUS: G-[SG]-[LFY]-x-R-[GE]-x(3)-[SGCL]-x-D-[LIVM]-D-[LIVMFY](3)-x(2)-[SAP].

NAME: Galactose-1-phosphate uridyl transferase family 1 active site signature.

CONSENSUS: F-E-N-[RK]-G-x(3)-G-x(4)-H-P-H-x-Q.

NAME: Galactose-1-phosphate uridyl transferase family 2 signature.

CONSENSUS: D-L-P-I-V-G-G-[ST]-[LIVM](2)-[SA]-H-[DEN]-H-[FY]-Q-G-G.

NAME: ADP-glucose pyrophosphorylase signature 1.

CONSENSUS: [AG]-G-G-x-G-[STK]-x-L-x(2)-L-[TA]-x(3)-A-x-P-A-[LV].

NAME: ADP-glucose pyrophosphorylase signature 2.

CONSENSUS: W-[FY]-x-G-[ST]-A-[DNSH]-[AS]-[LIVMFYW].

NAME: ADP-glucose pyrophosphorylase signature 3.

CONSENSUS: [APV]-[GS]-M-G-[LIVMN]-Y-[IVC]-[LIVMFY]-x(2)-[DENPHK].

NAME: Phosphatidate cytidyltransferase signature.

CONSENSUS: S-x-[LIVMF]-K-R-x(4)-K-D-x-[GSA]-x(2)-[LI]-[PG]-x-H-G-G-[LIVM]-x-D-R-

CONSENSUS: [LIVMFT]-D.

NAME: Ribonuclease PH signature.

CONSENSUS: C-[DE]-[LIVM](2)-Q-[GTA]-D-G-[SG]-x(2)-[TA]-A.

NAME: 2'-5'-oligoadenylate synthetases signature 1.

CONSENSUS: G-G-S-x-[AG]-[KR]-x-T-x-L-[KR]-[GST]-x-S-D-[AG].

NAME: 2'-5'-oligoadenylate synthetases signature 2.

CONSENSUS: R-P-V-I-L-D-P-x-[DE]-P-T.
NAME: CDP-alcohol phosphatidyltransferases signature.
CONSENSUS: D-G-x(2)-A-R-x(8)-G-x(3)-D-x(3)-D.
NAME: PEP-utilizing enzymes phosphorylation site signature.
CONSENSUS: G-[GA]-x-[TN]-x-H-[STA]-[STAV]-[LIVM](2)-[STAV]-[RG].
NAME: PEP-utilizing enzymes signature 2.
CONSENSUS: [DEQS]-x-[LIVMF]-S-[LIVMF]-G-[ST]-N-D-[LIVM]-x-Q-[LIVMFYGT]-[STALIV]-
CONSENSUS: [LIVMF]-[GAS]-x(2)-R.
NAME: Rhodanese signature 1.
CONSENSUS: [FY]-x(3)-H-[LIV]-P-G-A-x(2)-[LIVF].
NAME: Rhodanese C-terminal signature.
CONSENSUS: [AV]-x(2)-[FY]-[DEAP]-G-[GSA]-[WF]-x-E-[FYW].
NAME: CoA transferases signature 1.
CONSENSUS: [DN]-[GN]-x(2)-[LIVMFA](3)-G-G-F-x(3)-G-x-P.
NAME: CoA transferases signature 2.
CONSENSUS: [LF]-[HQ]-S-E-N-G-[LIVF](2)-[GA].
NAME: Phospholipase A2 histidine active site.
CONSENSUS: C-C-x(2)-H-x(2)-C.
NAME: Phospholipase A2 aspartic acid active site.
CONSENSUS: [LIVMA]-C-{[LIVMFYWPCST]}-C-D-x(5)-C.
NAME: Lipases, serine active site.
CONSENSUS: [LIV]-x-[LIVFY]-[LIVMST]-G-[HYWV]-S-x-G-[GSTAC].
NAME: Colipase signature.
CONSENSUS: Y-x(2)-Y-Y-x-C-x-C.
NAME: Lipolytic enzymes "G-D-S-L" family, serine active site.
CONSENSUS: [LIVMFYAG](4)-G-D-S-[LIVM]-x(1,2)-[TAG]-G.
NAME: Lipolytic enzymes "G-D-X-G" family, putative histidine active site.
CONSENSUS: [LIVMF](2)-x-[LIVMF]-H-G-G-[SAG]-[FY]-x(3)-[STDN]-x(2)-[ST]-H.
NAME: Lipolytic enzymes "G-D-X-G" family, putative serine active site.
CONSENSUS: [LIVM]-x-[LIVMF]-[SA]-G-D-S-[CA]-G-[GA]-x-L-[CA].
NAME: Carboxylesterases type-B serine active site.
CONSENSUS: F-[GR]-G-x(4)-[LIVM]-x-[LIV]-x-G-x-S-[STAG]-G.
NAME: Carboxylesterases type-B signature 2.
CONSENSUS: [ED]-D-C-L-[YT]-[LIV]-[DNS]-[LIV]-[LIVFYW]-x-[PQR].
NAME: Pectinesterase signature 1.
CONSENSUS: [GSTN]-x(5)-[LIVM]-x-[LIVM]-x(2)-G-x-Y-[DNK]-E-x-[LIVM]-x-[LIVM].
NAME: Pectinesterase signature 2.
CONSENSUS: G-[STAD]-[LIVMT]-D-F-I-F-G.
NAME: Peptidyl-tRNA hydrolase signature 1.
CONSENSUS: [FY]-x(2)-T-R-H-N-x-G-x(2)-[LIVMFA](2)-[DE].
NAME: Peptidyl-tRNA hydrolase signature 2.
CONSENSUS: [GS]-x(3)-H-N-G-[LIVM]-[KR]-[DNS]-[LIVMT].
NAME: Alkaline phosphatase active site.
CONSENSUS: [IV]-x-D-S-[GAS]-[GASC]-[GAST]-[GA]-T.
NAME: Histidine acid phosphatases phosphohistidine signature.
CONSENSUS: [LIVM]-x(2)-[LIVMA]-x(2)-[LIVM]-x-R-H-[GN]-x-R-x-[PAS].
NAME: Histidine acid phosphatases active site signature.
CONSENSUS: [LIVMF]-x-[LIVMFAG]-x(2)-[STAG]-H-D-[STANQ]-x-[LIVM]-x(2)-[LIVMFY]-x(2)-
CONSENSUS: [STA].
NAME: Class A bacterial acid phosphatases signature.

CONSENSUS: G-S-Y-P-S-G-H-T.

NAME: 5'-nucleotidase signature 1.

CONSENSUS: [LIVM]-x-[LIVM](2)-[HEA]-[TT]-x-D-x-H-[GSA]-x-[LIVMF].

NAME: 5'-nucleotidase signature 2.

CONSENSUS: [FYP]-x(4)-[LIVM]-G-N-H-E-F-[DN].

NAME: Fructose-1-6-bisphosphatase active site.

CONSENSUS: [AG]-[RK]-L-x(1,2)-[LIV]-[FY]-E-x(2)-P-[LIVM]-[GSA].

NAME: Serine/threonine specific protein phosphatases signature.

CONSENSUS: [LIVM]-R-G-N-H-E.

NAME: Protein phosphatase 2A regulatory subunit PR55 signature 1.

CONSENSUS: E-F-D-Y-L-K-S-L-E-I-E-E-K-I-N.

NAME: Protein phosphatase 2A regulatory subunit PR55 signature 2.

CONSENSUS: N-[AG]-H-[TA]-Y-H-I-N-S-I-S-[LIVM]-N-S-D.

NAME: Protein phosphatase 2C signature.

CONSENSUS: [LIVMFY]-[LIVMFYA]-[GSAC]-[LIVM]-[FYC]-D-G-H-[GAV].

NAME: Tyrosine specific protein phosphatases active site.

CONSENSUS: [LIVMF]-H-C-x(2)-G-x(3)-[STC]-[STAGP]-x-[LIVMFY].

NAME: Tyrosine specific protein phosphatases profile.

NAME: Dual specificity protein phosphatase profile.

NAME: PTP type protein phosphatase profile.

NAME: Inositol monophosphatase family signature 1.

CONSENSUS: [FWV]-x(0,1)-[LIVM]-D-P-[LIVM]-D-[SG]-[ST]-x(2)-[FY]-x-[HKRNSTY].

NAME: Inositol monophosphatase family signature 2.

CONSENSUS: [WV]-D-x-[AC]-[GSA]-[GSAPV]-x-[LIVACP]-[LIV]-[LIVAC]-x(3)-[GH]-[GA].

NAME: Prokaryotic zinc-dependent phospholipase C signature.

CONSENSUS: H-Y-x-[GT]-D-[LIVM]-[DNS]-x-P-x-H-[PA]-x-N.

NAME: Phosphatidylinositol-specific phospholipase X-box domain profile.

NAME: Phosphatidylinositol-specific phospholipase Y-box domain profile.

NAME: 3'5'-cyclic nucleotide phosphodiesterases signature.

CONSENSUS: H-D-[LIVMFY]-x-H-x-[AG]-x(2)-[NQ]-x-[LIVMFY].

NAME: cAMP phosphodiesterases class-II signature.

CONSENSUS: H-x-H-L-D-H-[LIVM]-x-[GS]-[LIVMA]-[LIVM](2)-x-S-[AP].

NAME: Sulfatases signature 1.

CONSENSUS: [SAP]-[LIVMST]-[CS]-[STAC]-P-[STA]-R-x(2)-[LIVMFW](2)-[TR]-G.

NAME: Sulfatases signature 2.

CONSENSUS: G-[YV]-x-[ST]-x(2)-[IVA]-G-K-x(0,1)-[FYWK]-[HL].

NAME: AP endonucleases family 1 signature 1.

CONSENSUS: [APF]-D-[LIVMF](2)-x-[LIVM]-Q-E-x-K.

NAME: AP endonucleases family 1 signature 2.

CONSENSUS: D-[ST]-[FY]-R-[KH]-x(7,8)-[FYW]-[ST]-[FYW](2).

NAME: AP endonucleases family 1 signature 3.

CONSENSUS: N-x-G-x-R-[LIVM]-D-[LIVMFYH]-x-[LV]-x-S.

NAME: AP endonucleases family 2 signature 1.

CONSENSUS: H-x(2)-Y-[LIVMF]-[IM]-N-[LIVMCA]-[AG].

NAME: AP endonucleases family 2 signature 2.

CONSENSUS: [GR]-[LIVMF]-C-[LIVM]-D-T-C-H.

NAME: AP endonucleases family 2 signature 3.

CONSENSUS: [LIVMW]-H-x-N-[DE]-[SA]-K-x(3)-G-[SA]-x(2)-D.

NAME: Deoxyribonuclease I signature 1.

CONSENSUS: [LIVM](2)-[AP]-L-H-[STA](2)-P-x(5)-E-[LIVM]-[DN]-x-L-x-[DE]-V.

NAME: Deoxyribonuclease I signature 2.

CONSENSUS: G-D-F-N-A-x-C-[SA].

NAME: Endonuclease III iron-sulfur binding region signature.

CONSENSUS: C-x(3)-[KRS]-P-[KRAGL]-C-x(2)-C-x(5)-C.

NAME: Endonuclease III family signature.

CONSENSUS: {GST}-x-[LIVMF]-P-x(5)-{LIVMW}-x(2,3)-[LT]-[PAS]-G-V-[GA]-x(3)-[GAC]-

CONSENSUS: x(3)-[LIVM]-x(2)-[SALV]-[LIVMFYW]-[GANK].

NAME: Ribonuclease II family signature.

CONSENSUS: [HI]-[FYE]-[GSTAM]-[LIVM]-x(4,5)-Y-[STAL]-x-[FWVAC]-[TV]-[SA]-P-[LIVMA]-

CONSENSUS: [RQ]-[KR]-[FY]-x-D-x(3)-[HQ].

NAME: Ribonuclease III family signature.

CONSENSUS: [DEQ]-[RQ]-[LM]-E-[FYW]-[LV]-G-D-[SAR].

NAME: Bacterial Ribonuclease P protein component signature.

CONSENSUS: [LIVMFYS]-x(2)-A-x(2)-R-[NH]-[KRQL]-[LIVM]-[KRA]-R-x-[LIVMTA]-[KR].

NAME: Ribonuclease T2 family histidine active site 1.

CONSENSUS: [FYWL]-x-[LIVM]-H-G-L-W-P.

NAME: Ribonuclease T2 family histidine active site 2.

CONSENSUS: [LIVMF]-x(2)-[HDGTY]-[EQ]-[FYW]-x-[KR]-H-G-x-C.

NAME: Pancreatic ribonuclease family signature.

CONSENSUS: C-K-x(2)-N-T-F.

NAME: DNA/RNA non-specific endonucleases active site.

CONSENSUS: D-R-G-H-[QIL]-x(3)-A.

NAME: Thermonuclease family signature 1.

CONSENSUS: D-G-D-T-[LIVM]-x-[LIVMC]-x(9,10)-R-[LIVM]-x(2)-[LIVM]-D-x-P-E.

NAME: Thermonuclease family signature 2.

CONSENSUS: D-[KR]-Y-[GQ]-R-x-[LV]-[GA]-x-[IV]-[FYW].

NAME: Beta-amylase active site 1.

CONSENSUS: H-x-C-G-G-N-V-G-D.

NAME: Beta-amylase active site 2.

CONSENSUS: G-x-[SA]-G-E-[LIVM]-R-Y-P-S-Y.

NAME: Glucoamylase active site region signature.

CONSENSUS: [STN]-[GP]-x(1,2)-[DE]-x-W-E-E-x(2)-[GS].

NAME: Polygalacturonase active site.

CONSENSUS: [GSDENKRH]-x(2)-[VMFC]-x(2)-[GS]-H-G-[LIVMAG]-x(1,2)-[LIVM]-G-S.

NAME: Clostridium cellulosome enzymes repeated domain signature.

CONSENSUS: D-[LIVMFY]-[DNV]-x-[DNS]-x(2)-[LIVM]-[DN]-[SALM]-x-D-x(3)-[LIVMF]-x-

CONSENSUS: [RKS]-x-[LIVMF].

NAME: Chitinases family 18 active site.

CONSENSUS: [LIVMFY]-[DN]-G-[LIVMF]-[DN]-[LIVMF]-[DN]-x-E.

NAME: Chitinases family 19 signature 1.

CONSENSUS: C-x(4,5)-F-Y-[ST]-x(3)-[FY]-[LIVMF]-x-A-x(3)-[YF]-x(2)-F-[GSA].

NAME: Chitinases family 19 signature 2.

CONSENSUS: [LIVM]-[GSA]-F-x-[STAG](2)-[LIVMFY]-W-[FY]-W-[LIVM].

NAME: Alpha-lactalbumin / lysozyme C signature.

CONSENSUS: C-x(3)-C-x(2)-[LMF]-x(3)-[DEN]-[LI]-x(5)-C.

NAME: Alpha-galactosidase signature.

CONSENSUS: G-[LIVMFY]-x(2)-[LIVMFY]-x-[LIVM]-D-D-x-W-x(3,4)-R-[DNSF].

NAME: Trehalase signature 1.

CONSENSUS: P-G-G-R-F-x-E-x-Y-x-W-D-x-Y.
NAME: Trehalase signature 2.
CONSENSUS: Q-W-D-x-P-x-[GA]-W-[PA]-P.
NAME: Alpha-L-fucosidase putative active site.
CONSENSUS: P-x(2)-L-x(3)-K-W-E-x-C.
NAME: Glycosyl hydrolases family 1 active site.
CONSENSUS: [LIVMFSTC]-[LIVFYS]-[LIV]-[LIVMST]-E-N-G-[LIVMFAR]-[CSAGN].
NAME: Glycosyl hydrolases family 1 N-terminal signature.
CONSENSUS: F-x-[FYWM]-[GSTA]-x-[GSTA]-x-[GSTA](2)-[FYNH]-[NQ]-x-E-x-[GSTA].
NAME: Glycosyl hydrolases family 2 signature 1.
CONSENSUS: N-x-[LIVMFYWD]-R-[STACN](2)-H-Y-P-x(4)-[LIVMFYW](2)-x(3)-[DN]-x(2)-
CONSENSUS: G-[LIVMFYW](4).
NAME: Glycosyl hydrolases family 2 acid/base catalyst.
CONSENSUS: [DENQF]-[KRVW]-N-H-[AP]-[SAC]-[LIVMF](3)-W-[GS]-x(2,3)-N-E.
NAME: Glycosyl hydrolases family 3 active site.
CONSENSUS: [LIVM](2)-[KR]-x-[EQK]-x(4)-G-[LIVMFT]-[LIVT]-[LIVMF]-[ST]-D-x(2)-
CONSENSUS: [SGADNI].
NAME: Glycosyl hydrolases family 5 signature.
CONSENSUS: [LIV]-[LIVMFYWGA](2)-[DNEQG]-[LIVMGST]-x-N-E-[PV]-[RHDNSTLIVFY].
NAME: Glycosyl hydrolases family 6 signature 1.
CONSENSUS: V-x-Y-x(2)-P-x-R-D-C-[GSAF]-x(2)-[GSA](2)-x-G.
NAME: Glycosyl hydrolases family 6 signature 2.
CONSENSUS: [LIVMYA]-[LIVA]-[LIVT]-[LIV]-E-P-D-[SAL]-[LI]-[PSAG].
NAME: Glycosyl hydrolases family 8 signature.
CONSENSUS: A-[ST]-D-[AG]-D-x(2)-[IM]-A-x-[SA]-[LIVM]-[LIVMG]-x-A-x(3)-[FW].
NAME: Glycosyl hydrolases family 9 active sites signature 1.
CONSENSUS: [STV]-x-[LIVMFY]-[STV]-x(2)-G-x-[NKR]-x(4)-[PLIVM]-H-x-R.
NAME: Glycosyl hydrolases family 9 active sites signature 2.
CONSENSUS: [FYW]-x-D-x(4)-[FYW]-x(3)-E-x-[STA]-x(3)-N-[STA].
NAME: Glycosyl hydrolases family 10 active site.
CONSENSUS: [GTA]-x(2)-[LIVN]-x-[IVMF]-[ST]-E-[LIY]-[DN]-[LIVMF].
NAME: Glycosyl hydrolases family 11 active site signature 1.
CONSENSUS: [PSA]-[LQ]-x-E-Y-Y-[LIVM](2)-[DE]-x-[FYWHN].
NAME: Glycosyl hydrolases family 11 active site signature 2.
CONSENSUS: [LIVMF]-x(2)-E-[AG]-[YWG]-[QRFGS]-[SG]-[STAN]-G-x-[SAF].
NAME: Glycosyl hydrolases family 16 active sites.
CONSENSUS: E-[LIV]-D-[LIV]-x(0,1)-E-x(2)-[GQ]-[KRNF]-x-[PSTA].
NAME: Glycosyl hydrolases family 17 signature.
CONSENSUS: [LIVM]-x-[LIVMFYWA](3)-[STAG]-E-[STA]-G-W-P-[STN]-x-[SAGQ].
NAME: Glycosyl hydrolases family 25 active sites signature.
CONSENSUS: D-[LIVM]-x(3)-[NQ]-[PG]-x(9,10)-G-x(4)-[LIVMFY](2)-K-x-[ST]-E-[GS]-x(2)-
CONSENSUS: Y-x-[DN].
NAME: Glycosyl hydrolases family 31 active site.
CONSENSUS: [GF]-[LIVMF]-W-x-D-M-[NSA]-E.
NAME: Glycosyl hydrolases family 31 signature 2.
CONSENSUS: G-[AV]-D-[LIVMT]-C-G-[FY]-x(3)-[ST]-x(3)-L-C-x-R-W-x(2)-[LV]-[GS]-[SA]-
CONSENSUS: F-x-P-F-x-R-[DN].
NAME: Glycosyl hydrolases family 32 active site.
CONSENSUS: H-x(2)-P-x(4)-[LIVM]-N-D-P-N-G.
NAME: Glycosyl hydrolases family 35 putative active site.
CONSENSUS: G-G-P-[LIVM](2)-x(2)-Q-x-E-N-E-[FY].

NAME: Glycosyl hydrolases family 39 active site.
 CONSENSUS: W-x-F-E-x-W-N-E-P-[DN].

NAME: Glycosyl hydrolases family 45 active site.
 CONSENSUS: [STA]-T-R-Y-[FYW]-D-x(5)-[CA].

NAME: Prokaryotic transglycosylases signature.
 CONSENSUS: [LIVM]-x(3)-E-S-x(3)-[AP]-x(3)-S-x(5)-G-[LIVM]-[LIVMFYW]-x-[LIVMFYW]-
 CONSENSUS: x(4)-[SAG].

NAME: Inosine-uridine preferring nucleoside hydrolase family signature.
 CONSENSUS: D-x-D-[PT]-[GA]-x-D-D-[TAV]-[VI]-A.

NAME: Alkylbase DNA glycosidases alkA family signature.
 CONSENSUS: G-I-G-x-W-[ST]-[AV]-x-[LIVMFY](2)-x-[LIVM]-x(8)-[MF]-x(2)-[ED]-D.

NAME: Formamidopyrimidine-DNA glycosylase signature.
 CONSENSUS: C-x(2,4)-C-x-[GTAQ]-x-[IV]-x(7)-R-[GSTAN]-[STA]-x-[FYI]-C-x(2)-C-Q.

NAME: Uracil-DNA glycosylase signature.
 CONSENSUS: [KR]-[LIV]-[LIVC]-[LIVM]-x-G-[QI]-D-P-Y.

NAME: S-adenosyl-L-homocysteine hydrolase signature 1.
 CONSENSUS: [CS]-N-x-[FYL]-S-[ST]-[QA]-[DEN]-x-[AV](2)-A-A-[LIV]-[SAV].

NAME: S-adenosyl-L-homocysteine hydrolase signature 2.
 CONSENSUS: G-K-x(3)-[LIV]-x-G-Y-G-x-V-G-[KR]-G-x-A.

NAME: Cytosol aminopeptidase signature.
 CONSENSUS: N-T-D-A-E-G-R-L.

NAME: Aminopeptidase P and proline dipeptidase signature.
 CONSENSUS: [HA]-[GSYR]-[LIVMT]-[SG]-H-x-[LIV]-G-[LIVM]-x-[IV]-H-[DE].

NAME: Methionine aminopeptidase subfamily 1 signature.
 CONSENSUS: [MFY]-x-G-H-G-[LIVMC]-[GSH]-x(3)-H-x(4)-[LIVM]-x-[HN]-[YWV].

NAME: Methionine aminopeptidase subfamily 2 signature.
 CONSENSUS: [DA]-[LIVMY]-x-K-[LIVM]-D-x-G-x-[HQ]-[LIVM]-[DNS]-G-x(3)-[DN].

NAME: Renal dipeptidase active site.
 CONSENSUS: [LIVM]-E-G-[GA]-x(2)-[LIVMF]-x(6)-L-x(3)-Y-x(2)-G-[LIVM]-R.

NAME: Serine carboxypeptidases, serine active site.
 CONSENSUS: [LIVM]-x-[GTA]-E-S-Y-[AG]-[GS].

NAME: Serine carboxypeptidases, histidine active site.
 CONSENSUS: [LIVF]-x(2)-[LIVSTA]-x-[IVPST]-x-[GSDNQL]-[SAGV]-[SG]-H-x-[IVAQ]-P-x(3)-
 CONSENSUS: [PSA].

NAME: Zinc carboxypeptidases, zinc-binding region 1 signature.
 CONSENSUS: [PK]-x-[LIVMFY]-x-[LIVMFY]-x(4)-H-[STAG]-x-E-x-[LIVM]-[STAG]-x(6)-
 CONSENSUS: [LIVMFYTA].

NAME: Zinc carboxypeptidases, zinc-binding region 2 signature.
 CONSENSUS: H-[STAG]-x(3)-[LIVME]-x(2)-[LIVMFYW]-P-[FYW].

NAME: Serine proteases, trypsin family, histidine active site.
 CONSENSUS: [LIVM]-[ST]-A-[STAG]-H-C.

NAME: Serine proteases, trypsin family, serine active site.
 CONSENSUS: [DNSTAGC]-[GSTAPIMVQH]-x(2)-G-[DE]-S-G-[GS]-[SAPHV]-[LIVMFYWH]-
 CONSENSUS: [LIVMFYSTANQH].

NAME: Serine proteases, subtilase family, aspartic acid active site.
 CONSENSUS: [STAIV]-x-[LIVMF]-[LIVM]-D-[DSTA]-G-[LIVMFC]-x(2,3)-[DNH].

NAME: Serine proteases, subtilase family, histidine active site.
 CONSENSUS: H-G-[STM]-x-[VIC]-[STAGC]-[GS]-x-[LIVMA]-[STAGCLV]-[SAGM].

NAME: Serine proteases, subtilase family, serine active site.
 CONSENSUS: G-T-S-x-[SA]-x-P-x(2)-[STAVC]-[AG].

NAME: Serine proteases, V8 family, histidine active site.
 CONSENSUS: [ST]-G-[LIVMFYW](3)-[GN]-x(2)-T-[LIVM]-x-T-x(2)-H.

NAME: Serine proteases, V8 family, serine active site.
 CONSENSUS: T-x(2)-[GC]-[NQ]-S-G-S-x-[LIVM]-[FY].

NAME: Serine proteases, ompin family signature 1.
 CONSENSUS: W-T-D-x-S-x-H-P-x-T.

NAME: Serine proteases, ompin family signature 2.
 CONSENSUS: A-G-Y-Q-E-[ST]-R-[FYW]-S-[FYW]-[TN]-A-x-G-G-[ST]-Y.

NAME: Prolyl endopeptidase family serine active site.
 CONSENSUS: D-x(3)-A-x(3)-[LIVMFYW]-x(14)-G-x-S-x-G-G-[LIVMFYW](2).

NAME: Endopeptidase Clp serine active site.
 CONSENSUS: T-x(2)-[LIVMF]-G-x-A-[SAC]-S-[MSA]-[PAG]-[STA].

NAME: Endopeptidase Clp histidine active site.
 CONSENSUS: R-x(3)-[EAP]-x(3)-[LIVMFYT]-M-[LIVM]-H-Q-P.

NAME: ATP-dependent serine proteases, lon family, serine active site.
 CONSENSUS: D-G-[PD]-S-A-[GS]-[LIVMCA]-[TA]-[LIVM].

NAME: Eukaryotic thiol (cysteine) proteases cysteine active site.
 CONSENSUS: Q-x(3)-[GE]-x-C-[YW]-x(2)-[STAGC]-[STAGCV].

NAME: Eukaryotic thiol (cysteine) proteases histidine active site.
 CONSENSUS: [LIVMGSTAN]-x-H-[GSACE]-[LIVM]-x-[LIVMAT](2)-G-x-[GSADNH].

NAME: Eukaryotic thiol (cysteine) proteases asparagine active site.
 CONSENSUS: [FYCH]-[WT]-[LIVT]-x-[KRQAG]-N-[ST]-W-x(3)-[FYW]-G-x(2)-G-[LFYW]-
 CONSENSUS: [LIVMFYG]-x-[LIVMF].

NAME: Ubiquitin carboxyl-terminal hydrolase family 1 cysteine active-site.
 CONSENSUS: Q-x(3)-N-[SA]-C-G-x(3)-[LIVM](2)-H-[SA]-[LIVM]-[SA].

NAME: Ubiquitin carboxyl-terminal hydrolases family 2 signature 1.
 CONSENSUS: G-[LIVMFY]-x(1,3)-[AGC]-[NASM]-x-C-[FYW]-[LIVMC]-[NST]-[SACV]-x-[LIVMS]-
 CONSENSUS: Q.

NAME: Ubiquitin carboxyl-terminal hydrolases family 2 signature 2.
 CONSENSUS: Y-x-L-x-[SAG]-[LIVMFT]-x(2)-H-x-G-x(4,5)-G-H-Y.

NAME: Caspase family histidine active site.
 CONSENSUS: H-x(2,4)-[SC]-x(4)-[LIVMF](2)-[ST]-H-G.

NAME: Caspase family cysteine active site.
 CONSENSUS: K-P-K-[LIVMF](4)-Q-A-C-[RQG]-G.

NAME: Eukaryotic and viral aspartyl proteases active site.
 CONSENSUS: [LIVMFGAC]-[LIVMTADN]-[LIVFSA]-D-[ST]-G-[STAV]-[STAPDENQ]-x-[LIVMFSTNC]-
 CONSENSUS: x-[LIVMFGTA].

NAME: Neutral zinc metallopeptidases, zinc-binding region signature.
 CONSENSUS: [GSTALIVN]-x(2)-H-E-[LIVMFYW]-[DEHRKP]-H-x-[LIVMFYWGSPQ].

NAME: Matrixins cysteine switch.
 CONSENSUS: P-R-C-[GN]-x-P-[DR]-[LIVSAPKQ].

NAME: Insulinase family, zinc-binding region signature.
 CONSENSUS: G-x(8,9)-G-x-[STA]-H-[LIVMFY]-[LIVMC]-[DERN]-[HRKL]-[LMFAT]-x-[LFSTH]-x-
 CONSENSUS: [GSTAN]-[GST].

//

AC PS01016;
 DE Glycoprotease family signature.
 CONSENSUS: [KR]-[GSAT]-x(4)-[FYWHL]-[DQNGK]-x-P-x-[LIVMFY]-x(3)-H-x(2)-[AG]-H-
 CONSENSUS: [LIVM].

NAME: Proteasome A-type subunits signature.
 CONSENSUS: [FY]-x(4)-[STNV]-x-[FYW]-S-P-x-G-[RKH]-x(2)-Q-[LIVM]-[DE]-Y-[SAD]-x(2)-
 CONSENSUS: [SAG].

NAME: Proteasome B-type subunits signature.

CONSENSUS: [LIVMA]-[GSA]-[LIVMF]-x-[FYLVGAC]-x(2)-[GSACFY]-[LIVMSTAC](3)-[GAC]-

CONSENSUS: [GSTACV]-[DES]-x(15)-[RK]-x(12,13)-G-x(2)-[GSTA]-D.

NAME: Signal peptidases I serine active site.

CONSENSUS: [GS]-x-S-M-x-[PS]-[AT]-[LF].

NAME: Signal peptidases I lysine active site.

CONSENSUS: K-R-[LIVMSTA](2)-G-x-[PG]-G-[DE]-x-[LIVM]-x-[LIVMFY].

NAME: Signal peptidases I signature 3.

CONSENSUS: [LIVMFYW](2)-x(2)-G-D-[NH]-x(3)-[SND]-x(2)-[SG].

NAME: Signal peptidases II signature.

CONSENSUS: [GAF]-[GA]-[GAS]-[LIVM]-[GAS]-N-[LVMFG]-[LIVMFY]-D-R-[LIMFA].

NAME: Peptidase family U32 signature.

CONSENSUS: E-x-F-x(2)-G-[SA]-[LIVM]-C-x(4)-G-x-C-x-[LIVM]-S.

NAME: Amidases signature.

CONSENSUS: G-[GA]-S-S-[GS]-G-x-[GSA]-[GSAVY]-x-[LIVM]-[GSA]-x(6)-[GSA]-x-[GA]-x-D-

CONSENSUS: x-[GA]-x-S-[LIVM]-R-x-P-[GSAC].

NAME: Asparaginase / glutaminase active site signature 1.

CONSENSUS: [LIVM]-x(2)-T-G-G-T-[TV]-[AGS].

NAME: Asparaginase / glutaminase active site signature 2.

CONSENSUS: G-x-[LIVM]-x(2)-H-G-T-D-T-[LIVM].

NAME: Urease nickel ligands signature.

CONSENSUS: T-[AY]-[GA]-[GAT]-[LIVM]-D-x-H-[LIVM]-H-x(3)-P.

NAME: Urease active site.

CONSENSUS: [LIVM](2)-[CT]-H-[HN]-L-x(3)-[LIVM]-x(2)-D-[LIVM]-x-F-A.

NAME: ArgE / dapE / ACY1 / CPG2 / yscS family signature 1.

CONSENSUS: [LIV]-[GALMY]-[LIVMF]-x-[GSA]-H-x-D-[TV]-[STAV].

NAME: ArgE / dapE / ACY1 / CPG2 / yscS family signature 2.

CONSENSUS: [GSTAI]-[SANQ]-D-x-K-[GSACN]-x(2)-[LIVMA]-x(2)-[LIVMFY]-x(14,17)-[LIVM]-

CONSENSUS: x-[LIVMF]-[LIVMSTAG]-[LIVMFA]-x(2)-[DNG]-E-E-x-[GSTN].

NAME: Dihydroorotase signature 1.

CONSENSUS: D-[LIVMFYWSAP]-H-[LIVA]-H-[LIVF]-[RN]-x-[PGN].

NAME: Dihydroorotase signature 2.

CONSENSUS: [GA]-[ST]-D-x-A-P-H-x(4)-K.

NAME: Beta-lactamase class-A active site.

CONSENSUS: [FY]-x-[LIVMFY]-x-S-[TV]-x-K-x(4)-[AGLM]-x(2)-[LC].

NAME: Beta-lactamase class-C active site.

CONSENSUS: F-E-[LIVM]-G-S-[LIVMG]-[SA]-K.

NAME: Beta-lactamase class-D active site.

CONSENSUS: [PA]-x-S-[ST]-F-K-[LIV]-[PAL]-x-[STA]-[LI].

NAME: Beta-lactamases class B signature 1.

CONSENSUS: [LI]-x-[STN]-[HN]-x-H-[GSTA]-D-x(2)-G-[GP]-x(7,8)-[GS].

NAME: Beta-lactamases class B signature 2.

CONSENSUS: P-x(3)-[LIVM](2)-x-G-x-C-[LIVMF](2)-K.

NAME: Arginase family signature 1.

CONSENSUS: [LIVMF]-G-G-x-H-x-[LIVMT]-[STAV]-x-[PAG]-x(3)-[GSTA].

NAME: Arginase family signature 2.

CONSENSUS: [LIVM](2)-x-[LIVMFY]-D-[AS]-H-x-D.

NAME: Arginase family signature 3.

CONSENSUS: [ST]-[LIVMFY]-D-[LIVM]-D-x(3)-[PAQ]-x(3)-P-[GSA]-x(7)-G.

NAME: Adenosine and AMP deaminase signature.

CONSENSUS: [SA]-[LIVM]-[NGS]-[STA]-D-D-P.
 NAME: Cytidine and deoxycytidylate deaminases zinc-binding region signature.
 CONSENSUS: [CH]-[AGV]-E-x(2)-[LIVMFGAT]-[LIVM]-x(17,33)-P-C-x(2,8)-C-x(3)-[LIVM].
 NAME: GTP cyclohydrolase I signature 1.
 CONSENSUS: [EN]-[LIVM](2)-x(2)-[KRQN]-[DN]-[LIVM]-x(3)-[ST]-x-C-E-H-H.
 NAME: GTP cyclohydrolase I signature 2.
 CONSENSUS: [SA]-x-[RK]-x-Q-[LIVM]-Q-E-[RN]-[LI]-[TSN].
 NAME: Nitrilases / cyanide hydratase signature 1.
 CONSENSUS: G-x(2)-[LIVMFY](2)-x-[IF]-x-E-x(2)-[LIVM]-x-G-Y-P.
 NAME: Nitrilases / cyanide hydratase active site signature.
 CONSENSUS: G-[GAQ]-x(2)-C-[WA]-E-[NH]-x(2)-[PST]-[LIVMFYS]-x-[KR].
 NAME: Inorganic pyrophosphatase signature.
 CONSENSUS: D-[SGDN]-D-[PE]-[LIVMF]-D-[LIVMGAC].
 NAME: Acylphosphatase signature 1.
 CONSENSUS: [LIV]-x-G-x-V-Q-G-V-x-[FM]-R.
 NAME: Acylphosphatase signature 2.
 CONSENSUS: G-[FYW]-[AVC]-[KRQAM]-N-x(3)-G-x-V-x(5)-G.
 NAME: ATP synthase alpha and beta subunits signature.
 CONSENSUS: P-[SAP]-[LIV]-[DNH]-x(3)-S-x-S.
 NAME: ATP synthase gamma subunit signature.
 CONSENSUS: [IV]-T-x-E-x(2)-[DE]-x(3)-G-A-x-[SAKR].
 NAME: ATP synthase delta (OSCP) subunit signature.
 CONSENSUS: [LIVM]-x-[LIVMFYT]-x(3)-[LIVMT]-[DENQK]-x(2)-[LIVM]-x-[GSA]-G-[LIVMFYGA]-
 CONSENSUS: x-[LIVM]-[KRHENQ]-x-[GSEN].
 NAME: ATP synthase a subunit signature.
 CONSENSUS: [STAGN]-x-[STAG]-[LIVMF]-R-L-x-[SAGV]-N-[LIVMT].
 NAME: ATP synthase c subunit signature.
 CONSENSUS: [GSTA]-R-[NQ]-P-x(10)-[LIVMFYW](2)-x(3)-[LIVMFYW]-x-[DE].
 NAME: E1-E2 ATPases phosphorylation site.
 CONSENSUS: D-K-T-G-T-[LI]-[TI].
 NAME: Sodium and potassium ATPases beta subunits signature 1.
 CONSENSUS: [FYW]-x(2)-[FYW]-x-[FYW]-[DN]-x(6)-[LIVM]-G-R-T-x(3)-W.
 NAME: Sodium and potassium ATPases beta subunits signature 2.
 CONSENSUS: [RK]-x(2)-C-[RKQWI]-x(5)-L-x(2)-C-[SA]-G.
 NAME: GDA1/CD39 family of nucleoside phosphatases signature.
 CONSENSUS: [LIVM]-x-G-x(2)-E-G-x-[FY]-x-[FW]-[LIVA]-[TAG]-x-N-[HY].
 NAME: Iodothyronine deiodinases active site.
 CONSENSUS: R-P-L-V-x-N-F-G-S-[CA]-T-C-P-x-F.
 NAME: Cutinase, serine active site.
 CONSENSUS: P-x-[STA]-x-[LIV]-[IVT]-x-[GS]-G-Y-S-[QL]-G.
 NAME: Cutinase, aspartate and histidine active sites.
 CONSENSUS: C-x(3)-D-x-[IV]-C-x-G-[GST]-x(2)-[LIVM]-x(2,3)-H.
 NAME: DDC / GAD / HDC / TyrDC pyridoxal-phosphate attachment site.
 CONSENSUS: S-[LIVMFYW]-x(5)-K-[LIVMFYWG](2)-x(3)-[LIVMFYW]-x-[CA]-x(2)-[LIVMFYWQ]-
 CONSENSUS: x(2)-[RK].
 NAME: Orn/Lys/Arg decarboxylases family 1 pyridoxal-P attachment site.
 CONSENSUS: [STAV]-x-S-x-H-K-x(2)-[GSTAN](2)-x-[STA]-Q-[STA](2).
 NAME: Orn/DAP/Arg decarboxylases family 2 pyridoxal-P attachment site.
 CONSENSUS: [FY]-[PA]-x-K-[SACV]-[NHCLFW]-x(4)-[LIVMF]-[LIVMTA]-x(2)-[LIVMA]-x(3)-
 CONSENSUS: [GTE].

NAME: Orn/DAP/Arg decarboxylases family 2 signature 2.
 CONSENSUS: [GS]-x(2,6)-[LIVMSCP]-x(2)-[LIVMF]-[DNS]-[LIVMCA]-G-G-G-[LIVMFY]-
 CONSENSUS: [GSTPCEQ].

NAME: Orotidine 5'-phosphate decarboxylase active site.
 CONSENSUS: [LIVMFTA]-[LIVMF]-x-D-x-K-x(2)-D-I-[GP]-x-T-[LIVMTA].

NAME: Phosphoenolpyruvate carboxylase active site 1.
 CONSENSUS: [VT]-x-T-A-H-P-T-[EQ]-x(2)-R-[KRH].

NAME: Phosphoenolpyruvate carboxylase active site 2.
 CONSENSUS: [IV]-M-[LIVM]-G-Y-S-D-S-x-K-D-[STAG]-G.

NAME: Phosphoenolpyruvate carboxykinase (GTP) signature.
 CONSENSUS: F-P-S-A-C-G-K-T-N.

NAME: Phosphoenolpyruvate carboxykinase (ATP) signature.
 CONSENSUS: L-I-G-D-D-E-H-x-W-x-[DE]-x-G-[IV]-x-N.

NAME: Uroporphyrinogen decarboxylase signature 1.
 CONSENSUS: P-x-W-x-M-R-Q-A-G-R.

NAME: Uroporphyrinogen decarboxylase signature 2.
 CONSENSUS: G-F-[STAGCV]-[STAGC]-x-P-[FYW]-T-[LV]-x(2)-Y-x(2)-[AE]-[GK].

NAME: Indole-3-glycerol phosphate synthase signature.
 CONSENSUS: [LIVMFY]-[LIVMC]-x-E-[LIVMFYC]-K-[KRSP]-[STAK]-S-P-[ST]-x(3)-[LIVMFYST].

NAME: Ribulose biphosphate carboxylase large chain active site.
 CONSENSUS: G-x-[DN]-F-x-K-x-D-E.

NAME: Fructose-bisphosphate aldolase class-I active site.
 CONSENSUS: [LIVM]-x-[LIVMFYW]-E-G-x-[LS]-L-K-P-[SN].

NAME: Fructose-bisphosphate aldolase class-II signature 1.
 CONSENSUS: [FYVM]-x(1,3)-[LIVMH]-[APN]-[LIVM]-x(1,2)-[LIVM]-H-x-D-H-[GACH].

NAME: Fructose-bisphosphate aldolase class-II signature 2.
 CONSENSUS: [LIVM]-E-x-E-[LIVM]-G-x(2)-[GM]-[GSTA]-x-E.

NAME: Malate synthase signature.
 CONSENSUS: [KR]-[DENQ]-H-x(2)-G-L-N-x-G-x-W-D-Y-[LIVM]-F.

NAME: Hydroxymethylglutaryl-coenzyme A lyase active site.
 CONSENSUS: S-V-A-G-L-G-G-C-P-Y.

NAME: Hydroxymethylglutaryl-coenzyme A synthase active site.
 CONSENSUS: N-x-[DN]-[IV]-E-G-[IV]-D-x(2)-N-A-C-[FY]-x-G.

NAME: Citrate synthase signature.
 CONSENSUS: G-[FYA]-[GA]-H-x-[IV]-x(1,2)-[RKT]-x(2)-D-[PS]-R.

NAME: Alpha-isopropylmalate and homocitrate synthases signature 1.
 CONSENSUS: L-R-[DE]-G-x-Q-x(10)-K.

NAME: Alpha-isopropylmalate and homocitrate synthases signature 2.
 CONSENSUS: [LIVMFW]-x(2)-H-x-H-[DN]-D-x-G-x-[GAS]-x-[GASLI].

NAME: KDPG and KHG aldolases active site.
 CONSENSUS: G-[LIVM]-x(3)-E-[LIV]-T-[LF]-R.

NAME: KDPG and KHG aldolases Schiff-base forming residue.
 CONSENSUS: G-x(3)-[LIVMF]-K-[LF]-F-P-[SA]-x(3)-G.

NAME: Isocitrate lyase signature.
 CONSENSUS: K-[KR]-C-G-H-[LMQ].

NAME: Beta-eliminating lyases pyridoxal-phosphate attachment site.
 CONSENSUS: Y-x-D-x(3)-M-S-[GA]-K-K-D-x-[LIVM](2)-x-[LIVM]-G-G.

NAME: DNA photolyases class 1 signature 1.
 CONSENSUS: T-G-x-P-[LIVM](2)-D-A-x-M-[RA]-x-[LIVM].

NAME: DNA photolyases class 1 signature 2.

CONSENSUS: [DN]-R-x-R-[LIVM](2)-x-[STA](2)-F-[LIVMFA]-x-K-x-L-x(2,3)-W-[KRQ].
NAME: DNA photolyases class 2 signature 1.
CONSENSUS: F-x-E-E-x-[LIVM](2)-R-R-E-L-x(2)-N-F.
NAME: DNA photolyases class 2 signature 2.
CONSENSUS: G-x-H-D-x(2)-W-x-E-R-x-[LIVM]-F-G-K-[LIVM]-R-[FY]-M-N.
NAME: Eukaryotic-type carbonic anhydrases signature.
CONSENSUS: S-E-H-x-[LIVM]-x(4)-[FYH]-x(2)-E-[LIVM]-H-[LIVMFA](2).
NAME: Prokaryotic-type carbonic anhydrases signature 1.
CONSENSUS: C-[SA]-D-S-R-[LIVM]-x-[AP].
NAME: Prokaryotic-type carbonic anhydrases signature 2.
CONSENSUS: [EQ]-Y-A-[LIVM]-x(2)-[LIVM]-x(4)-[LIVMF](3)-x-G-H-x(2)-C-G.
NAME: Fumarate lyases signature.
CONSENSUS: G-S-x(2)-M-x(2)-K-x-N.
NAME: Aconitase family signature 1.
CONSENSUS: [LIVM]-x(2)-[GSACIVM]-x-[LIV]-[GTIV]-[STP]-C-x(0,1)-T-N-[GSTANI]-x(4)-
CONSENSUS: [LIVMA].
NAME: Aconitase family signature 2.
CONSENSUS: G-x(2)-[LIVWPQ]-x(3)-[GAC]-C-[GSTAM]-[LIMPTA]-C-[LIMV]-[GA].
NAME: Dihydroxy-acid and 6-phosphogluconate dehydratases signature 1.
CONSENSUS: C-D-K-x(2)-P-[GA]-x(3)-[GA].
NAME: Dihydroxy-acid and 6-phosphogluconate dehydratases signature 2.
CONSENSUS: [SA]-L-[LIVM]-T-D-[GA]-R-[LIVMF]-S-[GA]-[GAV]-[ST].
NAME: Dehydroquinase class I active site.
CONSENSUS: D-[LIVM]-[DE]-[LIVN]-x(18,20)-[LIVM](2)-x-[SC]-[NHY]-H-[DN].
NAME: Dehydroquinase class II signature.
CONSENSUS: [LIVM]-[NQ]-G-P-N-[LV]-x(2)-L-G-x-R-[QED]-P-x(2)-[FY]-G.
NAME: Enolase signature.
CONSENSUS: [LIV](3)-K-x-N-Q-I-G-[ST]-[LIV]-[ST]-[DE]-[STA].
NAME: Serine/threonine dehydratases pyridoxal-phosphate attachment site.
CONSENSUS: [DESH]-x(4,5)-[STVG]-x-[AS]-[FYI]-K-[DLIFSA]-[RVMF]-[GA]-[LIVMGA].
NAME: Enoyl-CoA hydratase/isomerase signature.
CONSENSUS: [LIVM]-[STA]-x-[LIVM]-[DENQRHSTA]-G-x(3)-[AG](3)-x(4)-[LIVMST]-x-[CSTA]-
CONSENSUS: [DQHP]-[LIVMFY].
NAME: Imidazoleglycerol-phosphate dehydratase signature 1.
CONSENSUS: [LIVMY]-[DE]-x-H-H-x(2)-E-x(2)-[GCA]-[LIVM]-[STAC]-[LIVM].
NAME: Imidazoleglycerol-phosphate dehydratase signature 2.
CONSENSUS: G-x-[DN]-x-H-H-x(2)-E-[STAGC]-x-[FY]-K.
NAME: Tryptophan synthase alpha chain signature.
CONSENSUS: [LIVM]-E-[LIVM]-G-x(2)-[FYC]-[ST]-[DE]-[PA]-[LIVMY]-[AGLI]-[DE]-G.
NAME: Tryptophan synthase beta chain pyridoxal-phosphate attachment site.
CONSENSUS: [LIVM]-x-H-x-G-[STA]-H-K-x-N.
NAME: Delta-aminolevulinic acid dehydratase active site.
CONSENSUS: G-x-D-x-[LIVM](2)-[IV]-K-P-[GSA]-x(2)-Y.
NAME: Urocanase active site.
CONSENSUS: F-Q-G-L-P-x-R-I-C-W.
NAME: Prephenate dehydratase signature 1.
CONSENSUS: [FY]-x-[LIVM]-x(2)-[LIVM]-x(5)-[DN]-x(5)-T-R-F-[LIVMW]-x-[LIVM].
NAME: Prephenate dehydratase signature 2.
CONSENSUS: [LIVM]-[ST]-[KR]-[LIVM]-E-[ST]-R-P.
NAME: Dihydropicolinate synthetase signature 1.

CONSENSUS: [GSA]-[LIVM]-[LIVMFY]-x(2)-G-[ST]-[TG]-G-E-[GASNF]-x(6)-[EQ].
NAME: Dihydropicolinate synthetase signature 2.
CONSENSUS: Y-[DNS]-[LIVMF]-P-x(2)-[ST]-x(3)-[LIVM]-x(13,14)-[LIVM]-x-[SGA]-[LIVMF]-
CONSENSUS: K-[DEQAF]-[STAC].
NAME: RsuA family of pseudouridine synthase signature.
CONSENSUS: G-R-L-D-x(2)-[ST]-x-G-[LIVMF](4)-[ST]-[DNT].
NAME: Cysteine synthase/cystathionine beta-synthase P-phosphate attachment site.
CONSENSUS: K-x-E-x(3)-[PA]-[STAGC]-x-S-[IVAP]-K-x-R-x-[STAG]-x(2)-[LIVM].
NAME: Phenylalanine and histidine ammonia-lyases signature.
CONSENSUS: G-[STG]-[LIVM]-[STG]-[AC]-S-G-[DH]-L-x-P-L-[SA]-x(2)-[SA].
NAME: Porphobilinogen deaminase cofactor-binding site.
CONSENSUS: E-R-x-[LIVMFA]-x(3)-[LIVMF]-x-G-[GSA]-C-x-[IVT]-P-[LIVMF]-[GSA].
NAME: Cys/Met metabolism enzymes pyridoxal-phosphate attachment site.
CONSENSUS: [DQ]-[LIVMF]-x(3)-[STAGC]-[STAGCI]-T-K-[FYWQ]-[LIVMF]-x-G-[HQ]-[SGNH].
NAME: Glyoxalase I signature 1.
CONSENSUS: [HQ]-[IVT]-x-[LIVFY]-x-[IV]-x(5)-[STA]-x(2)-F-[YM]-x(2,3)-[LMF]-G-[LMF].
NAME: Glyoxalase I signature 2.
CONSENSUS: G-[NTKQ]-x(0,5)-[GA]-[LVFY]-[GH]-H-[IVF]-[CGA]-x-[STAGL]-x(2)-[DNC].
NAME: Cytochrome c and c1 heme lyases signature 1.
CONSENSUS: H-N-x(2)-N-E-x(2)-W-[NQKR]-x(4)-W-E.
NAME: Cytochrome c and c1 heme lyases signature 2.
CONSENSUS: P-F-D-R-H-D-W.
NAME: Adenylate cyclases class-I signature 1.
CONSENSUS: E-Y-F-G-[SA](2)-L-W-x-L-Y-K.
NAME: Adenylate cyclases class-I signature 2.
CONSENSUS: Y-R-N-x-W-[NS]-E-[LIVM]-R-T-L-H-F-x-G.
NAME: Guanylate cyclases signature.
CONSENSUS: G-V-[LIVM]-x(0,1)-G-x(5)-[FY]-x-[LIVM]-[FYW]-[GS]-[DNTHKW]-[DNT]-[IV]-
CONSENSUS: [DNTA]-x(5)-[DE].
NAME: Chorismate synthase signature 1.
CONSENSUS: G-E-S-H-[GC]-x(2)-[LIVM]-[GTV]-x-[LIVM](2)-[DE]-G-x-[PV].
NAME: Chorismate synthase signature 2.
CONSENSUS: [GE]-R-[SA](2)-[SAG]-R-[EV]-[ST]-x(2)-[RH]-V-x(2)-G.
NAME: Chorismate synthase signature 3.
CONSENSUS: R-[SH]-D-[PSV]-[CSAV]-x(4)-[GAI]-x-[IVGSP]-[LIVM]-x-E-[STAH]-[LIVM].
NAME: 6-pyruvoyl tetrahydropterin synthase signature 1.
CONSENSUS: C-N-N-x(2)-G-H-G-H-N-Y.
NAME: 6-pyruvoyl tetrahydropterin synthase signature 2.
CONSENSUS: D-H-K-N-L-D-x-D.
NAME: Ferrochelatase signature.
CONSENSUS: [LIVMF](2)-x-S-x-H-[GS]-[LIVM]-P-x(4,5)-[DENQKR]-x-G-D-x-Y.
NAME: Alanine racemase pyridoxal-phosphate attachment site.
CONSENSUS: V-x-K-A-[DN]-[GA]-Y-G-H-G.
NAME: Aspartate and glutamate racemases signature 1.
CONSENSUS: [IVA]-[LIVM]-x-C-x(0,1)-N-[ST]-[MSA]-[STH]-[LIVFYSTANK].
NAME: Aspartate and glutamate racemases signature 2.
CONSENSUS: [LIVM](2)-x-[AG]-C-T-[DEH]-[LIVMFY]-[PNGRS]-x-[LIVM].
NAME: Mandelate racemase / muconate lactonizing enzyme family signature 1.
CONSENSUS: A-x-[SAG](2)-[LIVM]-[DE]-x-A-x(2)-D-x(2)-[GA]-[KR].
NAME: Mandelate racemase / muconate lactonizing enzyme family signature 2.

CONSENSUS: G-x(7)-D-x(9)-A-x(14)-[LIVM]-E-[DENQ]-P-x(4)-[DENQ].
NAME: Ribulose-phosphate 3-epimerase family signature 1.
CONSENSUS: [LIVMF]-H-[LIVMFY]-D-[LIVM]-x-D-x(1,2)-[FY]-[LIVM]-x-N-x-[STAV].
NAME: Ribulose-phosphate 3-epimerase family signature 2.
CONSENSUS: [LIVMA]-x-[LIVM]-M-[ST]-[VS]-x-P-x(3)-G-Q-x-F-x(6)-[NK]-[LIVMC].
NAME: Aldose 1-epimerase putative active site.
CONSENSUS: [NS]-x-T-N-H-x-Y-[FW]-N-[LI].
NAME: Cyclophilin-type peptidyl-prolyl cis-trans isomerase signature.
CONSENSUS: [FY]-x(2)-[STCNLV]-x-F-H-[RH]-[LIVMN]-[LIVM]-x(2)-F-[LIVM]-x-Q-[AG]-G.
NAME: Cyclophilin-type peptidyl-prolyl cis-trans isomerase profile.
NAME: FKBP-type peptidyl-prolyl cis-trans isomerase signature 1.
CONSENSUS: [LIVMC]-x-[YF]-x-[GVL]-x(1,2)-[LFT]-x(2)-G-x(3)-[DE]-[STAEQK]-[STAN].
NAME: FKBP-type peptidyl-prolyl cis-trans isomerase signature 2.
CONSENSUS: [LIVMFY]-x(2)-[GA]-x(3,4)-[LIVMF]-x(2)-[LIVMFHK]-x(2)-G-x(4)-[LIVMF]-
CONSENSUS: x(3)-[PSGAQ]-x(2)-[AG]-[FY]-G.
NAME: FKBP-type peptidyl-prolyl cis-trans isomerase domain profile.
NAME: PpiC-type peptidyl-prolyl cis-trans isomerase signature.
CONSENSUS: F-[GSADEI]-x-[LVAQ]-A-x(3)-[ST]-x(3,4)-[STQ]-x(3,5)-[GER]-G-x-[LIVM]-
CONSENSUS: [GS].
NAME: Triosephosphate isomerase active site.
CONSENSUS: [AV]-Y-E-P-[LIVM]-W-[SA]-I-G-T-[GK].
NAME: Xylose isomerase signature 1.
CONSENSUS: [LI]-E-P-K-P-x(2)-P.
NAME: Xylose isomerase signature 2.
CONSENSUS: [FL]-H-D-x-D-[LIV]-x-[PD]-x-[GDE].
NAME: Phosphomannose isomerase type I signature 1.
CONSENSUS: Y-x-D-x-N-H-K-P-E.
NAME: Phosphomannose isomerase type I signature 2.
CONSENSUS: H-A-Y-[LIVM]-x-G-x(2)-[LIVM]-E-x-M-A-x-S-D-N-x-[LIVM]-R-A-G-x-T-P-K.
NAME: Phosphoglucose isomerase signature 1.
CONSENSUS: [DENS]-x-[LIVM]-G-G-R-[FY]-S-[LIVMT]-x-[STA]-[PSAC]-[LIVMA]-G.
NAME: Phosphoglucose isomerase signature 2.
CONSENSUS: [GS]-x-[LIVM]-[LIVMFYW]-x(4)-[FY]-[DN]-Q-x-G-V-E-x(2)-K.
NAME: Glucosamine/galactosamine-6-phosphate isomerases signature.
CONSENSUS: [LIVM]-x(3)-G-x-[LIT]-x-[LIV]-x-[LIVM]-x-G-[LIVM]-G-x-[DEN]-G-H.
NAME: Phosphoglycerate mutase family phosphohistidine signature.
CONSENSUS: [LIVM]-x-R-H-G-[EQ]-x(3)-N.
NAME: Phosphoglucomutase and phosphomannomutase phosphoserine signature.
CONSENSUS: [GSA]-[LIVM]-x-[LIVM]-[ST]-[PGA]-S-H-x-P-x(4)-[GNHE].
NAME: Methylmalonyl-CoA mutase signature.
CONSENSUS: R-I-A-R-N-[TQ]-x(2)-[LIVMFY](2)-x-[EQ]-E-x(4)-[KRN]-x(2)-D-P-x-[GSA]-
CONSENSUS: G-S.
NAME: Terpene synthases signature.
CONSENSUS: [DE]-G-S-W-x-G-x-W-[GA]-[LIVM]-x-[FY]-x-Y-[GA].
NAME: Eukaryotic DNA topoisomerase I active site.
CONSENSUS: [DEN]-x(6)-[GS]-[IT]-S-K-x(2)-Y-[LIVM]-x(3)-[LIVM].
NAME: Prokaryotic DNA topoisomerase I active site.
CONSENSUS: [EQ]-x-L-Y-[DEQT]-x(3,12)-[LI]-[ST]-Y-x-R-[ST]-[DEQS].
NAME: DNA topoisomerase II signature.
CONSENSUS: [LIVMA]-x-E-G-[DN]-S-A-x-[STAG].

NAME: Aminoacyl-transfer RNA synthetases class-I signature.

CONSENSUS: P-x(0,2)-[GSTAN]-[DENQGAPK]-x-[LIVMFP]-[HT]-[LIVMYAC]-G-[HNTG]-
CONSENSUS: [LIVMFYSTAGPC].

NAME: Aminoacyl-transfer RNA synthetases class-II signature 1.

CONSENSUS: [FYH]-R-x-[DE]-x(4,12)-[RH]-x(3)-F-x(3)-[DE].

NAME: Aminoacyl-transfer RNA synthetases class-II signature 2.

CONSENSUS: [GSTALVF]-[DENQHRKP]-[GSTA]-[LIVMF]-[DE]-R-[LIVMF]-x-[LIVMSTAG]-[LIVMFY].

NAME: WHEP-TRS domain signature.

CONSENSUS: [QY]-G-[DNEA]-x-[LIV]-[KR]-x(2)-K-x(2)-[KRNG]-[AS]-x(4)-[LIV]-[DENK]-
CONSENSUS: x(2)-[IV]-x(2)-L-x(3)-K.

NAME: ATP-citrate lyase / succinyl-CoA ligases family signature 1.

CONSENSUS: S-[KR]-S-G-[GT]-[LIVM]-[GST]-x-[EQ]-x(8,10)-G-x(4)-[LIVM]-[GA]-[LIVM]-G-
CONSENSUS: G-D.

NAME: ATP-citrate lyase / succinyl-CoA ligases family active site.

CONSENSUS: G-x(2)-A-x(4,7)-[RQT]-[LIVMF]-G-H-[AS]-[GH].

NAME: ATP-citrate lyase / succinyl-CoA ligases family signature 3.

CONSENSUS: G-x-[IV]-x(2)-[LIVMF]-x-[NA]-G-[GA]-G-[LA]-[STAV]-x(4)-D-x-[LIVM]-x(3)-
CONSENSUS: G-[GRE].

NAME: Glutamine synthetase signature 1.

CONSENSUS: [FYWL]-D-G-S-S-x(6,8)-[DENQSTAK]-[SA]-[DE]-x(2)-[LIVMFY].

NAME: Glutamine synthetase putative ATP-binding region signature.

CONSENSUS: K-P-[LIVMFYA]-x(3,5)-[NPAT]-G-[GSTAN]-G-x-H-x(3)-S.

NAME: Glutamine synthetase class-I adenylation site.

CONSENSUS: K-[LIVM]-x(5)-[LIVMA]-D-[RK]-[DN]-[LI]-Y.

NAME: D-alanine--D-alanine ligase signature 1.

CONSENSUS: H-G-x(2)-G-E-D-G-x-[LIVMA]-[QSA]-[GSA].

NAME: D-alanine--D-alanine ligase signature 2.

CONSENSUS: [LIV]-x(3)-[GA]-x-[GSAIV]-R-[LIVCA]-D-[LIVMF](2)-x(7,9)-[LI]-x-E-
CONSENSUS: [LIVA]-N-[STP]-x-P-[GA].

NAME: SAICAR synthetase signature 1.

CONSENSUS: [LIVMF](2)-P-[LIVM]-E-x-[LIVM]-[LIVMCA]-R-x(3)-[TA]-G-S.

NAME: SAICAR synthetase signature 2.

CONSENSUS: [LIVM]-[LIVMA]-D-x-K-[LIVMFY]-E-F-G.

NAME: Folylpolylglutamate synthase signature 1.

CONSENSUS: [LIVMFY]-x-[LIVM]-[STAG]-G-T-[NK]-G-K-x-[ST]-x(7)-[LIVM](2)-x(3)-[GSK].

NAME: Folylpolylglutamate synthase signature 2.

CONSENSUS: [LIVMFY](2)-E-x-G-[LIVM]-[GA]-G-x(2)-D-x-[GST]-x-[LIVM](2).

NAME: Ubiquitin-activating enzyme signature 1.

CONSENSUS: K-A-C-S-G-K-F-x-P.

NAME: Ubiquitin-activating enzyme active site.

CONSENSUS: P-[LIVM]-C-T-[LIVM]-[KRH]-x-[FT]-P.

NAME: Ubiquitin-conjugating enzymes active site.

CONSENSUS: [FYWLSP]-H-[PC]-[NH]-[LIV]-x(3,4)-G-x-[LIV]-C-[LIV]-x-[LIV].

NAME: Formate--tetrahydrofolate ligase signature 1.

CONSENSUS: G-[LIVM]-K-G-G-A-A-G-G-G-Y.

NAME: Formate--tetrahydrofolate ligase signature 2.

CONSENSUS: V-A-T-[IV]-R-A-L-K-x-[HN]-G-G.

NAME: Adenylosuccinate synthetase GTP-binding site.

CONSENSUS: Q-W-G-D-E-G-K-G.

NAME: Adenylosuccinate synthetase active site.

CONSENSUS: G-I-[GR]-P-x-Y-x(2)-K-x(2)-R.

NAME: Argininosuccinate synthase signature 1.
 CONSENSUS: A-[FY]-S-G-G-L-D-T-S.

NAME: Argininosuccinate synthase signature 2.
 CONSENSUS: G-x-T-x-K-G-N-D-x(2)-R-F.

NAME: Phosphoribosylglycinamide synthetase signature.
 CONSENSUS: R-F-G-D-P-E-x-[QM].

NAME: Carbamoyl-phosphate synthase subdomain signature 1.
 CONSENSUS: [FYV]-[PS]-[LIVMC]-[LIVMA]-[LIVM]-[KR]-[PSA]-[STA]-x(3)-[SG]-G-x-[AG].

NAME: Carbamoyl-phosphate synthase subdomain signature 2.
 CONSENSUS: [LIVMF]-[LIMN]-E-[LIVMCA]-N-[PATLIVM]-[KR]-[LIVMSTAC].

NAME: ATP-dependent DNA ligase AMP-binding site.
 CONSENSUS: [EDQH]-x-K-x-[DN]-G-x-R-[GACIVM].

NAME: ATP-dependent DNA ligase signature 2.
 CONSENSUS: E-G-[LIVMA]-[LIVM](2)-[KR]-x(5,8)-[YW]-[QNEK]-x(2,6)-[KRH]-x(3,5)-K-
 CONSENSUS: [LIVMFY]-K.

NAME: NAD-dependent DNA ligase signature 1.
 CONSENSUS: K-[LIVM]-D-G-[LIVM]-[SA]-x(4)-Y-x(2)-G-x-L-x(4)-[ST]-R-G-[DN]-G-x(2)-G-
 CONSENSUS: [DE]-[DENL].

NAME: NAD-dependent DNA ligase signature 2.
 CONSENSUS: [IV]-G-[KR]-[ST]-G-x-[LIVM]-[STNK]-x-[VT]-x(2)-L-x-[PS]-V.

NAME: RNA 3'-terminal phosphate cyclase signature.
 CONSENSUS: [RH]-G-x(2)-P-x-G(3)-x-[LIV].

NAME: Lipoate-protein ligase B signature.
 CONSENSUS: R-G-G-x(2)-T-[FYW]-H-x(2)-[GH]-Q-x-[LIV]-x-Y.

NAME: Isopenicillin N synthetase signature 1.
 CONSENSUS: [RK]-x-[STA]-x(2)-S-x-C-Y-[SL].

NAME: Isopenicillin N synthetase signature 2.
 CONSENSUS: [LIVM](2)-x-C-G-[STA]-x(2)-[STAG]-x(2)-T-x-[DNG].

NAME: Site-specific recombinases active site.
 CONSENSUS: Y-[LIVAC]-R-[VA]-S-[ST]-x(2)-Q.

NAME: Site-specific recombinases signature 2.
 CONSENSUS: G-[DE]-x(2)-[LIVM]-x(3)-[LIVM]-[DT]-R-[LIVM]-[GSA].

NAME: Transposases, Mutator family, signature.
 CONSENSUS: D-x(3)-G-[LIVMF]-x(6)-[STAV]-[LIVMFYW]-[PT]-x-[STAV]-x(2)-[QR]-x-C-x(2)-
 CONSENSUS: H.

NAME: Transposases, IS30 family, signature.
 CONSENSUS: R-G-x(2)-E-N-x-N-G-[LIVM](2)-R-[QE]-[LIVMFY](2)-P-K.

NAME: Autoinducers synthetases family signature.
 CONSENSUS: [LMFY]-R-x(3)-F-x(2)-[KR]-x(2)-W-x-[LIVM]-x(6,9)-E-x-D-x-[FY]-D.

NAME: Thiamine pyrophosphate enzymes signature.
 CONSENSUS: [LIVMF]-[GSA]-x(5)-P-x(4)-[LIVMFYW]-x-[LIVMF]-x-G-D-[GSA]-[GSAC].

NAME: Biotin-requiring enzymes attachment site.
 CONSENSUS: [GN]-[DEQTR]-x-[LIVMFY]-x(2)-[LIVM]-x-[AIV]-M-K-[LMAT]-x(3)-[LIVM]-x-
 CONSENSUS: [SAV].

NAME: 2-oxo acid dehydrogenases acyltransferase component lipoyl binding site.
 CONSENSUS: [GN]-x(2)-[LIVF]-x(5)-[LIVFC]-x(2)-[LIVFA]-x(3)-K-[STAIV]-[STAVQDN]-
 CONSENSUS: x(2)-[LIVMFS]-x(5)-[GCN]-x-[LIVMFY].

NAME: Putative AMP-binding domain signature.
 CONSENSUS: [LIVMFY]-x(2)-[STG]-[STAG]-G-[ST]-[STEI]-[SG]-x-[PASLIVM]-[KR].

NAME: Molybdenum cofactor biosynthesis proteins signature 1.
 CONSENSUS: [LIVM](3)-[LIT](2)-G-G-T-G-x(4)-D.

NAME: Molybdenum cofactor biosynthesis proteins signature 2.

CONSENSUS: S-x-[GS]-x(2)-D-x(5)-[LIVW]-x(10,12)-[LIV]-x(2)-[KR]-P-G-[KRL]-P-x(2)-[LIVMF]-[GA].

NAME: moaA / nifB / pqqE family signature.

CONSENSUS: [LIV]-x(3)-C-[NP]-[LIVMF]-[QRS]-C-x-[FYM]-C.

NAME: Radical activating enzymes signature.

CONSENSUS: [GV]-x-G-x-[KR]-x(3)-F-x(2)-G-x(0,1)-C-x(3)-C-x(2)-C-x-[NL].

NAME: Tpx family signature.

CONSENSUS: S-x-D-L-P-F-A-x(2)-[KR]-[FW]-C.

NAME: Cytochrome c family heme-binding site signature.

CONSENSUS: C-{CPWHF}-{CPWR}-C-H-{CFYW}.

NAME: Cytochrome b5 family, heme-binding domain signature.

CONSENSUS: [FY]-[LIVMK]-x(2)-H-P-[GA]-G.

NAME: Cytochrome b/b6 heme-ligand signature.

CONSENSUS: [DENQ]-x(3)-G-[FYWMQ]-x-[LIVMF]-R-x(2)-H.

NAME: Cytochrome b/b6 Qo site signature.

CONSENSUS: P-[DE]-W-[FY]-[LFY](2).

NAME: Cytochrome b559 subunits heme-binding site signature.

CONSENSUS: [LIV]-x-[ST]-[LIVF]-R-[FYW]-x(2)-[IV]-H-[STGA]-[LIV]-[STGA]-[IV]-P.

NAME: Nickel-dependent hydrogenases b-type cytochrome subunit signature 1.

CONSENSUS: R-[LIVMFYW]-x-H-W-[LIVM]-x(2)-[LIVMF]-[STAC]-[LIVM]-x(2)-L-x-[LIVM]-T-G.

NAME: Nickel-dependent hydrogenases b-type cytochrome subunit signature 2.

CONSENSUS: [RH]-[STA]-[LIVMFYW]-H-[RH]-[LIVM]-x(2)-W-x-[LIVMF]-x(2)-F-x(3)-H.

NAME: Succinate dehydrogenase cytochrome b subunit signature 1.

CONSENSUS: R-P-[LIVMT]-x(3)-[LIVM]-x(6)-[LIVMWPK]-x(4)-S-x(2)-H-R-x-[ST].

NAME: Succinate dehydrogenase cytochrome b subunit signature 2.

CONSENSUS: H-x(3)-[GA]-[LIVMT]-R-[HF]-[LIVMF]-x-[FYWM]-D-x-[GVA].

NAME: Thioredoxin family active site.

CONSENSUS: [LIVMF]-[LIVMSTA]-x-[LIVMFYC]-[FYWSTHE]-x(2)-[FYWGTM]-C-[GATPLVE]-[PHYWSTA]-C-x(6)-[LIVMFYWT].

NAME: Glutaredoxin active site.

CONSENSUS: [LIVD]-[FYSA]-x(4)-C-[PV]-[FYW]-C-x(2)-[TAV]-x(2,3)-[LIV].

NAME: Type-1 copper (blue) proteins signature.

CONSENSUS: [GA]-x(0,2)-[YSA]-x(0,1)-[VFY]-x-C-x(1,2)-[PG]-x(0,1)-H-x(2,4)-[MQ].

NAME: 2Fe-2S ferredoxins, iron-sulfur binding region signature.

CONSENSUS: C-{C}-{C}-[GA]-{C}-C-[GAST]-{CPDEKRHFYW}-C.

NAME: Adrenodoxin family, iron-sulfur binding region signature.

CONSENSUS: C-x(2)-[STAQ]-x-[STAMV]-C-[STA]-T-C-[HR].

NAME: 4Fe-4S ferredoxins, iron-sulfur binding region signature.

CONSENSUS: C-x(2)-C-x(2)-C-x(3)-C-[PEG].

NAME: High potential iron-sulfur proteins signature.

CONSENSUS: C-x(6,9)-[LIVM]-x(3)-G-[YW]-C-x(2)-[FYW].

NAME: Rieske iron-sulfur protein signature 1.

CONSENSUS: C-[TK]-H-L-G-C-[LIVT].

NAME: Rieske iron-sulfur protein signature 2.

CONSENSUS: C-P-C-H-x-[GSA].

NAME: Flavodoxin signature.

CONSENSUS: [LIV]-[LIVFY]-[FY]-x-[ST]-x(2)-[AGC]-x-T-x(3)-A-x(2)-[LIV].

NAME: Rubredoxin signature.

CONSENSUS: [LIVM]-x(3)-W-x-C-P-x-C-[AGD].

NAME: Electron transfer flavoprotein alpha-subunit signature.

CONSENSUS: [LI]-Y-[LIVM]-[AT]-x-G-[IV]-[SD]-G-x-[IV]-Q-H-x(2)-G-x(6)-[IV]-x-A-[IV]-N.

NAME: Electron transfer flavoprotein beta-subunit signature.

CONSENSUS: [IVA]-x-[KR]-x(2)-[DE]-[GD]-[GDE]-x(1,2)-[EQ]-x-[LIV]-x(4)-P-x-[LIVM](2)-[TAC].

NAME: Vertebrate metallothioneins signature.

CONSENSUS: C-x-C-[GSTAP]-x(2)-C-x-C-x(2)-C-x-C-x(2)-C-x-K.

NAME: Ferritin iron-binding regions signature 1.

CONSENSUS: E-x-[KR]-E-x(2)-E-[KR]-[LF]-[LIVMA]-x(2)-Q-N-x-R-x-G-R.

NAME: Ferritin iron-binding regions signature 2.

CONSENSUS: D-x(2)-[LIVMF]-[STAC]-[DH]-F-[LI]-[EN]-x(2)-[FY]-L-x(6)-[LIVM]-[KN].

NAME: Bacterioferritin signature.

CONSENSUS: <M-x-G-x(3)-V-[LIV]-x(2)-[LM]-x(3)-L-x(3)-L.

NAME: Transferrins signature 1.

CONSENSUS: Y-x(0,1)-[VAS]-V-[IVAC]-[IVA]-[IVA]-[RKH]-[RKS]-[GDENSA].

NAME: Transferrins signature 2.

CONSENSUS: Y-x-G-A-[FL]-[KRHNQ]-C-L-x(3,4)-G-[DENQ]-V-[GA]-[FYW].

NAME: Transferrins signature 3.

CONSENSUS: [DENQ]-[YF]-x-[LY]-L-C-x-[DN]-x(5,8)-[LIV]-x(4,5)-C-x(2)-A-x(4)-[HQR]-x-[LIVMFYW]-[LIVM].

NAME: Globins profile.

NAME: Protozoan/cyanobacterial globins signature.

CONSENSUS: F-[LF]-x(5)-G-[PA]-x(4)-G-[KRA]-x-[LIVM]-x(3)-H.

NAME: Plant hemoglobins signature.

CONSENSUS: [SN]-P-x-L-x(2)-H-A-x(3)-F.

NAME: Hemerythrins signature.

CONSENSUS: W-L-x-[NQ]-H-I-x(3)-D-F.

NAME: Arthropod hemocyanins / insect LSPs signature 1.

CONSENSUS: Y-[FYW]-x-E-D-[LIVM]-x(2)-N-x(6)-H-x(3)-P.

NAME: Arthropod hemocyanins / insect LSPs signature 2.

CONSENSUS: T-x(2)-R-D-P-x-[FY]-[FYW].

NAME: Heavy-metal-associated domain.

CONSENSUS: [LIVN]-x(2)-[LIVMFA]-x-C-x-[STAGCDNH]-C-x(3)-[LIVFG]-x(3)-[LIV]-x(9,11)-[IVA]-x-[LVFYS].

NAME: ABC transporters family signature.

CONSENSUS: [LIVMFYC]-[SA]-[SAPGLVFKQH]-G-[DENQMW]-[KRQASPLIMFW]-[KRNQSTAVM]-[KRACLVM]-[LIVMFYPAN]-[PHY]-[LIVMFW]-[SAGCLIVP]-[FYWHP]-[KRHP]-[LIVMFYWSTA].

NAME: Binding-protein-dependent transport systems inner membrane comp. sign.

CONSENSUS: [LIVMFY]-x(8)-[EQR]-[STAGV]-[STAG]-x(3)-G-[LIVMFYSTAC]-x(5)-[LIVMFYSTA]-x(4)-[LIVMFY]-[PKR].

NAME: ABC-2 type transport system integral membrane proteins signature.

CONSENSUS: [LIMST]-x(2)-[LIMW]-x(2)-[LIMCA]-[GSTC]-x-[GSAIV]-x(6)-[LIMGA]-[PGSNQ]-x(9,12)-P-[LIMFT]-x-[HRSY]-x(5)-[RQ].

NAME: Bacterial extracellular solute-binding proteins, family 1 signature.

CONSENSUS: [GAP]-[LIVMFA]-[STAVDN]-x(4)-[GSAV]-[LIVMFY](2)-Y-[ND]-x(3)-[LIVMF]-x-[KNDE].

NAME: Bacterial extracellular solute-binding proteins, family 3 signature.

CONSENSUS: G-[FYIL]-[DE]-[LIVMT]-[DE]-[LIVMF]-x(3)-[LIVMA]-[VAGC]-x(2)-[LIVMAGN].

NAME: Bacterial extracellular solute-binding proteins, family 5 signature.

CONSENSUS: [AG]-x(6,7)-[DNEG]-x(2)-[STAVE]-[LIVMFYWA]-x-[LIVMFY]-x-[LIVM]-[KR]-

CONSENSUS: [KRHDE]-[GDN]-[LIVMA]-[KNGSP]-[FW].
 NAME: Serum albumin family signature.
 CONSENSUS: [FY]-x(6)-C-C-x(7)-C-[LFY]-x(6)-[LIVMFYW].
 NAME: Transthyretin signature 1.
 CONSENSUS: S-K-C-P-L-M-V-K-V-L-D-[AS]-V-R-G.
 NAME: Transthyretin signature 2.
 CONSENSUS: S-P-[FY]-S-[FY]-S-T-T-A-[LIVM]-V-[ST]-x-P.
 NAME: Avidin / Streptavidin family signature.
 CONSENSUS: [DEN]-x(2)-[KR]-[STA]-x(2)-V-G-x-[DN]-x-[FW]-T-[KR].
 NAME: Eukaryotic cobalamin-binding proteins signature.
 CONSENSUS: [SN]-V-D-T-[GA]-A-[LIVM]-A-x-L-A-[LIVMF]-T-C.
 NAME: Lipocalin signature.
 CONSENSUS: [DENG]-x-[DENQGSTARK]-x(0,2)-[DENQARK]-[LIVFY]-[CP]-G-[C]-W-[FYWLRH]-x-
 CONSENSUS: [LIVMTA].
 NAME: Cytosolic fatty-acid binding proteins signature.
 CONSENSUS: [GSAIVK]-x-[FYW]-x-[LIVMF]-x(4)-[NHG]-[FY]-[DE]-x-[LIVMFY]-[LIVM]-x(2)-
 CONSENSUS: [LIVMAKR].
 NAME: Acyl-CoA-binding protein signature.
 CONSENSUS: P-[STA]-x-[DEN]-x-[LIVMF]-x(2)-[LIVMFY]-Y-[GSTA]-x-[FY]-K-Q-[STA](2)-x-G.
 NAME: LBP / BPI / CETP family signature.
 CONSENSUS: [PA]-[GA]-[LIVMC]-x(2)-R-[IV]-[ST]-x(3)-L-x(5)-[EQ]-x(4)-[LIVM]-[EQK]-
 CONSENSUS: x(8)-P.
 NAME: Phosphatidylethanolamine-binding protein family signature.
 CONSENSUS: [FY]-x-[LIVMF](3)-x-[DC]-P-D-x-P-[SN]-x(10)-H.
 NAME: Plant lipid transfer proteins signature.
 CONSENSUS: [LIVM]-[PA]-x(2)-C-x-[LIVM]-x-[LIVM]-x-[LIVMFY]-x-[LIVM]-[ST]-x(3)-
 CONSENSUS: [DN]-C-x(2)-[LIVM].
 NAME: Uteroglobin family signature 1.
 CONSENSUS: [GA]-x(3)-I-C-P-x-[LIVMF]-x(3)-[LIVM]-[DE]-x-[LIVMF](2).
 NAME: Uteroglobin family signature 2.
 CONSENSUS: [DEQ]-x(4)-[SN]-x(5)-[DEQ]-x-I-x(2)-S-[PSE]-[LS]-C.
 NAME: Mitochondrial energy transfer proteins signature.
 CONSENSUS: P-x-[DE]-x-[LIVAT]-[RK]-x-[LRH]-[LIVMFY]-[QMAIGV].
 NAME: Sugar transport proteins signature 1.
 CONSENSUS: [LIVMSTAG]-[LIVMFSAG]-x(2)-[LIVMSA]-[DE]-x-[LIVMFYWA]-G-R-[RK]-x(4,6)-
 CONSENSUS: [GSTA].
 NAME: Sugar transport proteins signature 2.
 CONSENSUS: [LIVMF]-x-G-[LIVMFA]-x(2)-G-x(8)-[LIFY]-x(2)-[EQ]-x(6)-[RK].
 NAME: LacY family proton/sugar symporters signature 1.
 CONSENSUS: G-[LIVM](2)-x-D-[RK]-L-G-L-[RK](2)-x-[LIVM](2)-W.
 NAME: LacY family proton/sugar symporters signature 2.
 CONSENSUS: P-x-[LIVMF](2)-N-R-[LIVM]-G-x-K-N-[STA]-[LIVM](3).
 NAME: PTR2 family proton/oligopeptide symporters signature 1.
 CONSENSUS: [GA]-[GAS]-[LIVMFYWA]-[LIVM]-[GAS]-D-x-[LIVMFYWT]-[LIVMFYW]-G-x(3)-[TAV]-
 CONSENSUS: [IV]-x(3)-[GSTAV]-x-[LIVMF]-x(3)-[GA].
 NAME: PTR2 family proton/oligopeptide symporters signature 2.
 CONSENSUS: [FYT]-x(2)-[LMFY]-[FYV]-[LIVMFYWA]-x-[IVG]-N-[LIVMAG]-G-[GSA]-[LIMF].
 NAME: Amiloride-sensitive sodium channels signature.
 CONSENSUS: Y-x(2)-[EQTF]-x-C-x(2)-[GSTDNL]-C-x-[QT]-x(2)-[LIVMT]-[LIVMS]-x(2)-C-x-C.
 NAME: Sodium:alanine symporter family signature.
 CONSENSUS: G-G-x-[GA](2)-[LIVM]-F-W-M-W-[LIVM]-x-[STAV]-[LIVMFA](2)-G.

NAME: Sodium:dicarboxylate symporter family signature 1.
 CONSENSUS: P-x(0,1)-G-[DE]-x-[LIVMF](2)-x-[LIVM](2)-[KREQ]-[LIVM](3)-x-P.

NAME: Sodium:dicarboxylate symporter family signature 2.
 CONSENSUS: P-x-G-x-[STA]-x-[NT]-[LIVMC]-D-G-[STAN]-x-[LIVM]-[FY]-x(2)-[LIVM]-x(2)-
 CONSENSUS: [LIVM]-[FY]-[LI]-[SA]-Q.

NAME: Sodium:galactoside symporter family signature.
 CONSENSUS: D-x(3)-G-x(3)-[DN]-x(6,8)-G-[KH]-F-[KR]-P-[FYW]-[LIVM](2)-x-[GSTA](2).

NAME: Sodium:neurotransmitter symporter family signature 1.
 CONSENSUS: W-R-F-[GP]-Y-x(4)-N-G-G-G-x-[FY].

NAME: Sodium:neurotransmitter symporter family signature 2.
 CONSENSUS: Y-[LIVMFY]-x(2)-[SC]-[LIVMFY]-[STQ]-x(2)-L-P-W-x(2)-C-x(4)-N-[GST].

NAME: Sodium:solute symporter family signature 1.
 CONSENSUS: [GS]-x(2)-[LIY]-x(3)-[LIVMFYWSTAG](10)-[LIY]-[TAV]-x(2)-G-G-[LMF]-x-
 CONSENSUS: [SAP].

NAME: Sodium:solute symporter family signature 2.
 CONSENSUS: [GAST]-[LIVM]-x(3)-[KR]-x(4)-G-A-x(2)-[GAS]-[LIVMGS]-[LIVMW]-[LIVMGAT]-G-
 CONSENSUS: x-[LIVMG].

NAME: Sodium:sulfate symporter family signature.
 CONSENSUS: [STACP]-S-x(2)-F-x(2)-P-[LIVM]-[GSA]-x(3)-N-x-[LIVM]-V.

NAME: glpT family of transporters signature.
 CONSENSUS: R-G-x(5)-W-N-x(2)-H-N-x-G-G.

NAME: Ammonium transporters signature.
 CONSENSUS: D-[FYWS]-A-G-[GSC]-x(2)-[IV]-x(3)-[SAG](2)-x(2)-[SAG]-[LIVMF]-x(3)-
 CONSENSUS: [LIVMFYWA](2)-x-[GK]-x-R.

NAME: BCCT family of transporters signature.
 CONSENSUS: [GSDN]-W-T-[LIVM]-x-[FY]-W-x-W-W.

NAME: Flagellar motor protein motA family signature.
 CONSENSUS: A-[LMF]-x-[GAT]-T-[LIVF]-x-G-x-[LIVMF]-x(7)-P.

NAME: Formate and nitrite transporters signature 1.
 CONSENSUS: [LIVMA]-[LIVMY]-x-G-[GSTA]-[DES]-L-[FI]-[TN]-[GS].

NAME: Formate and nitrite transporters signature 2.
 CONSENSUS: [GA]-x(2)-[CA]-N-[LIVMFYW](2)-V-C-[LV]-A.

NAME: Prokaryotic sulfate-binding proteins signature 1.
 CONSENSUS: K-x-[NQEK]-[GT]-G-[DQ]-x-[LIVM]-x(3)-Q-S.

NAME: Prokaryotic sulfate-binding proteins signature 2.
 CONSENSUS: N-P-K-[ST]-S-G-x-A-R.

NAME: Sulfate transporters signature.
 CONSENSUS: P-x-Y-[GS]-L-Y-[STAG](2)-x(4)-[LIVMFY](3)-x(3)-[GSTA](2)-S-[KR].

NAME: Amino acid permeases signature.
 CONSENSUS: [STAGC]-G-[PAG]-x(2,3)-[LIVMFYWA](2)-x-[LIVMFYW]-x-[LIVMFWSTAGC](2)-
 CONSENSUS: [STAGC]-x(3)-[LIVMFYW]-x-[LIVMST]-x(3)-[LIMCTA]-[GA]-E-x(5)-[PSAL].

NAME: Aromatic amino acids permeases signature.
 CONSENSUS: I-G-[GA]-G-M-[LF]-[SA]-x-P-x(3)-[SA]-G-x(2)-F.

NAME: Xanthine/uracil permeases family signature.
 CONSENSUS: [LIVM]-P-x-[PASIF]-V-[LIVM]-G-G-x(4)-[LIVM]-[FY]-[GSA]-x-[LIVM]-x(3)-G.

NAME: Anion exchangers family signature 1.
 CONSENSUS: F-G-G-[LIVM](2)-[KR]-D-[LIVM]-[RK]-R-R-Y.

NAME: Anion exchangers family signature 2.
 CONSENSUS: [FI]-L-I-S-L-I-F-I-Y-E-T-F-x-K-L.

NAME: MIP family signature.
 CONSENSUS: [HNQA]-x-N-P-[STA]-[LIVMF]-[ST]-[LIVMF]-[GSTAFY].

NAME: General diffusion Gram-negative porins signature.

CONSENSUS: [LIVMFY]-x(2)-G-x(2)-Y-x-F-x-K-x(2)-[SN]-[STAV]-[LIVMFYW]-V.

NAME: OmpA-like domain.

CONSENSUS: [LIVMA]-x-[GT]-x-[TA]-[DA]-x(2)-[DG]-[GSTP]-x(2)-[LFYDE]-[NQS]-x(2)-

CONSENSUS: [LI]-[SG]-[QE]-[KRQE]-R-A-x(2)-[LV]-x(3)-[LIVMF]-x(4,5)-[LIVM]-x(4)-

CONSENSUS: [LIVM]-x(3)-[SG]-x-G.

NAME: Eukaryotic mitochondrial porin signature.

CONSENSUS: [YH]-x(2)-D-[SPA]-x-[STA]-x(3)-[TAG]-[KR]-[LIVMF]-[DNSTA]-[DNS]-x(4)-

CONSENSUS: [GSTAN]-[LIVMA]-x-[LIVMY].

NAME: Insulin-like growth factor binding proteins signature.

CONSENSUS: G-C-[GS]-C-C-x(2)-C-A-x(6)-C.

NAME: GPR1/FUN34/yaaH family signature.

CONSENSUS: N-P-[AV]-P-[LF]-G-L-x-[GSA]-F.

NAME: GNS1/SUR4 family signature.

CONSENSUS: L-x-F-L-H-x-Y-H-H.

NAME: 43 Kd postsynaptic protein signature.

CONSENSUS: G-Q-D-Q-T-K-Q-Q-I.

NAME: Actins signature 1.

CONSENSUS: [FY]-[LIV]-G-[DE]-E-A-Q-x-[RKQ](2)-G.

NAME: Actins signature 2.

CONSENSUS: W-[IV]-[STA]-[RK]-x-[DE]-Y-[DNE]-[DE].

NAME: Actins and actin-related proteins signature.

CONSENSUS: [LM]-[LIVM]-T-E-[GAPQ]-x-[LIVMFYWHQ]-N-[PSTAQ]-x(2)-N-[KR].

NAME: Annexins repeated domain signature.

CONSENSUS: [TG]-[STV]-x(8)-[LIVMF]-x(2)-R-x(3)-[DEQNH]-x(7)-[IFY]-x(7)-[LIVMF]-

CONSENSUS: x(3)-[LIVMF]-x(11)-[LIVMFA]-x(2)-[LIVMF].

NAME: Caveolins signature.

CONSENSUS: F-E-D-V-I-A-E-P.

NAME: Clathrin light chain signature 1.

CONSENSUS: F-L-A-Q-Q-E-S.

NAME: Clathrin light chain signature 2.

CONSENSUS: [KR]-D-x-S-[KR]-[LIVM]-[KR]-x-[LIVM](3)-x-L-K.

NAME: Clusterin signature 1.

CONSENSUS: C-K-P-C-L-K-x-T-C.

NAME: Clusterin signature 2.

CONSENSUS: C-L-[RK]-M-[RK]-x-[EQ]-C-[ED]-K-C.

NAME: Connexins signature 1.

CONSENSUS: C-[DN]-T-x-Q-P-G-C-x(2)-V-C-Y-D.

NAME: Connexins signature 2.

CONSENSUS: C-x(3,4)-P-C-x(3)-[LIVM]-[DEN]-C-[FY]-[LIVM]-[SA]-[KR]-P.

NAME: Crystallins beta and gamma 'Greek key' motif signature.

CONSENSUS: [LIVMFYWA]-x-[DEHRKSTP]-[FY]-[DEQHKY]-x(3)-[FY]-x-G-x(4)-[LIVMFCST].

NAME: Dynamin family signature.

CONSENSUS: L-P-[RK]-G-[STN]-[GN]-[LIVM]-V-T-R.

NAME: Dynein light chain type 1 signature.

CONSENSUS: H-x-I-x-G-[KR]-x-F-[GA]-S-x-V-[ST]-[HY]-E.

NAME: FtsZ protein signature 1.

CONSENSUS: N-[ST]-D-x-Q-x-L-x(16,18)-G-x-G-[ATV]-G-[GSAN]-x-P-x(2)-G.

NAME: FtsZ protein signature 2.

CONSENSUS: [DNHKR]-[LIVMF]-x-[LIVMF](2)-[VSTAC]-[STAC]-G-x-G-[GK]-G-T-G-[ST]-G-

CONSENSUS: [GSAR]-[STA]-P-[LIVMFT]-[LIVMF]-[SGAV].

NAME: Fungal hydrophobins signature.

CONSENSUS: [GN]-[DNQPSA]-x-C-[GSTANK]-[GSTADNQ]-[STNQI]-[PTTV]-x-C-C-[DENQKPST].

NAME: Intermediate filaments signature.

CONSENSUS: [IV]-x-[TACI]-Y-[RKH]-x-[LM]-L-[DE].

NAME: Involucrin signature.

CONSENSUS: <M-S-[QH]-Q-x-T-[LV]-P-V-T-[LV].

NAME: Kinesin motor domain signature.

CONSENSUS: [GSA]-[KRHPSTQVM]-[LIVMF]-x-[LIVMF]-[IVC]-D-L-[AH]-G-[SAN]-E.

NAME: Kinesin motor domain profile.

NAME: Kinesin light chain repeat.

CONSENSUS: [DEQR]-A-L-x(3)-[GEQ]-x(3)-G-x-[DNS]-x-P-x-V-A-x(3)-N-x-L-[AS]-

CONSENSUS: x(5)-[QR]-x-[KR]-[FY]-x(2)-[AV]-x(4)-[HKNQ].

NAME: Myelin basic protein signature.

CONSENSUS: V-V-H-F-F-K-N.

NAME: Myelin P0 protein signature.

CONSENSUS: S-[KR]-S-x-K-[AG]-x-[SA]-E-K-K-[STA]-K.

NAME: Myelin proteolipid protein signature 1.

CONSENSUS: G-[MV]-A-L-F-C-G-C-G-H.

NAME: Myelin proteolipid protein signature 2.

CONSENSUS: C-x-[ST]-x-[DE]-x(3)-[ST]-[FY]-x-L-[FY]-I-x(4)-G-A.

NAME: Neuromodulin (GAP-43) signature 1.

CONSENSUS: <M-L-C-C-[LIVM]-R-R.

NAME: Neuromodulin (GAP-43) signature 2.

CONSENSUS: S-F-R-G-H-I-x-R-K-K-[LIVM].

NAME: Osteopontin signature.

CONSENSUS: [KQ]-x-[TA]-x(2)-[GA]-S-S-E-E-K.

NAME: Peripherin / rom-1 signature.

CONSENSUS: D-[GS]-V-P-F-[ST]-C-C-N-P-x-S-P-R-P-C.

NAME: Profilin signature.

CONSENSUS: <x(0,1)-[STA]-x(0,1)-W-[DENQH]-x-[YT]-x-[DEQ].

NAME: Surfactant associated polypeptide SP-C palmitoylation sites.

CONSENSUS: I-P-C-C-P-V.

NAME: Synapsins signature 1.

CONSENSUS: L-R-R-R-L-S-D-S.

NAME: Synapsins signature 2.

CONSENSUS: G-H-A-H-S-G-M-G-K-V-K.

NAME: Synaptobrevin signature.

CONSENSUS: N-[LIVM]-[DENS]-[KL]-V-x-[DEQ]-R-x(2)-[KR]-[LIVM]-[STDE]-x-[LIVM]-x-[DE]-

CONSENSUS: [KR]-[TA]-[DE].

NAME: Synaptophysin / synaptoporin signature.

CONSENSUS: L-S-V-[DE]-C-x-N-K-T.

NAME: Tropomyosins signature.

CONSENSUS: L-K-E-A-E-x-R-A-E.

NAME: Tubulin subunits alpha, beta, and gamma signature.

CONSENSUS: [SAG]-G-G-T-G-[SA]-G.

NAME: Tubulin-beta mRNA autoregulation signal.

CONSENSUS: <M-R-[DE]-[IL].

NAME: Tau and MAP proteins tubulin-binding domain signature.

CONSENSUS: G-S-x(2)-N-x(2)-H-x-[PA]-[AG]-G(2).

NAME: Neuraxin and MAP1B proteins repeated region signature.

CONSENSUS: [STAGDN]-Y-x-Y-E-x(2)-[DE]-[KR]-[STAGCI].
 NAME: F-actin capping protein alpha subunit signature 1.
 CONSENSUS: V-H-[FY](2)-E-D-G-N-V.
 NAME: F-actin capping protein alpha subunit signature 2.
 CONSENSUS: F-K-[AE]-L-R-R-x-L-P.
 NAME: F-actin capping protein beta subunit signature.
 CONSENSUS: C-D-Y-N-R-D.
 NAME: Vinculin family talin-binding region signature.
 CONSENSUS: [KR]-x-[LIVMF]-x(3)-[LIVMA]-x(2)-[LIVM]-x(6)-R-Q-Q-E-L.
 NAME: Vinculin repeated domain signature.
 CONSENSUS: [LIVM]-x-[QA]-A-x(2)-W-[IL]-x-[DN]-P.
 NAME: Amyloidogenic glycoprotein extracellular domain signature.
 CONSENSUS: G-[VT]-E-[FY]-V-C-C-P.
 NAME: Amyloidogenic glycoprotein intracellular domain signature.
 CONSENSUS: G-Y-E-N-P-T-Y-[KR].
 NAME: Cadherins extracellular repeated domain signature.
 CONSENSUS: [LIV]-x-[LIV]-x-D-x-N-D-[NH]-x-P.
 NAME: Insect cuticle proteins signature.
 CONSENSUS: G-x(7)-[DEN]-G-x(6)-Y-x-A-[DNG]-x(2,3)-G-[FY]-x-[AP].
 NAME: Gas vesicles protein GVPa signature 1.
 CONSENSUS: [LIVM]-x-[DE]-[LIVMFYT]-[LIVM]-[DE]-x-[LIVM](2)-[DKR](2)-G-x-[LIVM](2).
 NAME: Gas vesicles protein GVPa signature 2.
 CONSENSUS: R-[LIVA](3)-A-[GS]-[LIVMFY]-x-T-x(3)-Y-[AG].
 NAME: Gas vesicles protein GVPC repeated domain signature.
 CONSENSUS: F-L-x(2)-T-x(3)-R-x(3)-A-x(2)-Q-x(3)-L-x(2)-F.
 NAME: Bacterial microcompartments proteins signature.
 CONSENSUS: D-x(0,1)-M-x-K-[SAG](2)-x-[IV]-x-[LIVM]-[LIVMA]-[GCS]-x(4)-[GD]-[SGPD]-[GA].
 NAME: Flagella basal body rod proteins signature.
 CONSENSUS: [GTARYQ]-x(9)-[LIVMYSTA](2)-[GSTA]-[STADEN]-N-[LIVM]-[SAN]-N-x-[SADNFR]-[STV].
 NAME: Flagella transport protein fliP family signature 1.
 CONSENSUS: [PA]-A-[FY]-x-[LIVT]-[STH]-[EQ]-[LI]-x(2)-[GA]-F-[KREQ]-[IM]-G-[LIF].
 NAME: Flagella transport protein fliP family signature 2.
 CONSENSUS: P-[LIVMF]-K-[LIVMF](5)-x-[LIVMA]-[DNGS]-G-W.
 NAME: Plant viruses icosahedral capsid proteins 'S' region signature.
 CONSENSUS: [FYW]-x-[PSTA]-x(7)-G-x-[LIVM]-x-[LIVM]-x-[FYWT]-x(2)-D-x(5)-P.
 NAME: Potexviruses and carlaviruses coat protein signature.
 CONSENSUS: [RK]-[FYW]-A-[GAP]-F-D-x-F-x(2)-[LV]-x(3)-[GAST](2).
 NAME: Neurotransmitter-gated ion-channels signature.
 CONSENSUS: C-x-[LIVMFQ]-x-[LIVMF]-x(2)-[FY]-P-x-D-x(3)-C.
 NAME: ATP P2X receptors signature.
 CONSENSUS: G-G-x-[LIVM]-G-[LIVM]-x-[IV]-x-W-x-C-[DN]-L-D-x(5)-C-x-P-x-Y-x-F.
 NAME: G-protein coupled receptors signature.
 CONSENSUS: [GSTALIVMFYWC]-[GSTANCPDE]-[EDPKRH]-x(2)-[LIVMNQGA]-x(2)-[LIVMFT]-[GSTANC]-[LIVMFYWSTAC]-[DENH]-R-[FYWCSH]-x(2)-[LIVM].
 NAME: G-protein coupled receptors family 2 signature 1.
 CONSENSUS: C-x(3)-[FYWLIV]-D-x(3,4)-C-[FW]-x(2)-[STAGV]-x(8,9)-C-[PF].
 NAME: G-protein coupled receptors family 2 signature 2.
 CONSENSUS: Q-G-[LMFCA]-[LIVMFT]-[LIV]-x-[LIVFST]-[LIF]-[VFYH]-C-[LFY]-x-N-x(2)-V.

NAME: G-protein coupled receptors family 3 signature 1.
 CONSENSUS: [LV]-x-N-[LIVM](2)-x-L-F-x-I-[PA]-Q-[LIVM]-[STA]-x-[STA](3)-[STAN].

NAME: G-protein coupled receptors family 3 signature 2.
 CONSENSUS: C-C-[FYW]-x-C-x(2)-C-x(4)-[FYW]-x(2,4)-[DN]-x(2)-[STAH]-C-x(2)-C.

NAME: G-protein coupled receptors family 3 signature 3.
 CONSENSUS: F-N-E-[STA]-K-x-I-[STAG]-F-[ST]-M.

NAME: Visual pigments (opsins) retinal binding site.
 CONSENSUS: [LIVMWAC]-[PGAC]-x(3)-[SAC]-K-[STALIMR]-[GSACPNV]-[STACP]-x(2)-[DENF]-
 CONSENSUS: [AP]-x(2)-[IY].

NAME: Bacterial rhodopsins signature 1.
 CONSENSUS: R-Y-x-[DT]-W-x-[LIVMF]-[ST]-T-P-[LIVM](3).

NAME: Bacterial rhodopsins retinal binding site.
 CONSENSUS: [FYIV]-x-[FYVG]-[LIVM]-D-[LIVMF]-x-[STA]-K-x(2)-[FY].

NAME: Receptor tyrosine kinase class II signature.
 CONSENSUS: [DN]-[LIV]-Y-x(3)-Y-Y-R.

NAME: Receptor tyrosine kinase class III signature.
 CONSENSUS: G-x-H-x-N-[LIVM]-V-N-L-L-G-A-C-T.

NAME: Receptor tyrosine kinase class V signature 1.
 CONSENSUS: F-x-[DN]-x-[GAW]-[GA]-C-[LIVM]-[SA]-[LIVM](2)-[SA]-[LV]-[KRHQ]-[LIVA]-
 CONSENSUS: x(3)-[KR]-C-[PSAW].

NAME: Receptor tyrosine kinase class V signature 2.
 CONSENSUS: C-x(2)-[DE]-G-[DEQ]-W-x(2,3)-[PAQ]-[LIVMT]-[GT]-x-C-x-C-x(2)-G-[HFY]-
 CONSENSUS: [EQ].

NAME: Growth factor and cytokines receptors family signature 1.
 CONSENSUS: C-[LVFYR]-x(7,8)-[STVDN]-C-x-W.

NAME: Growth factor and cytokines receptors family signature 2.
 CONSENSUS: [STGL]-x-W-[SG]-x-W-S.

NAME: TNFR/NGFR family cysteine-rich region signature.
 CONSENSUS: C-x(4,6)-[FYH]-x(5,10)-C-x(0,2)-C-x(2,3)-C-x(7,11)-C-x(4,6)-[DNEQSKP]-
 CONSENSUS: x(2)-C.

NAME: TNFR/NGFR family cysteine-rich region domain.

NAME: Integrins alpha chain signature.
 CONSENSUS: [FYWS]-[RK]-x-G-F-F-x-R.

NAME: Integrins beta chain cysteine-rich domain signature.
 CONSENSUS: C-x-[GNQ]-x(1,3)-G-x-C-x-C-x(2)-C-x-C.

NAME: Natriuretic peptides receptors signature.
 CONSENSUS: G-P-x-C-x-Y-x-A-A-x-V-x-R-x(3)-H-W.

NAME: Photosynthetic reaction center proteins signature.
 CONSENSUS: [NH]-x(4)-P-x-H-x(2)-[SAG]-x(11)-[SAGC]-x-H-[SAG](2).

NAME: Antenna complexes alpha subunits signature.
 CONSENSUS: [LIVFAG]-x-[GASV]-[LIVFA]-x-[IV]-H-x(3)-[LIVM]-[GSTAE]-[STANH]-x(1,3)-
 CONSENSUS: [STN]-W-[LIVMFYW].

NAME: Antenna complexes beta subunits signature.
 CONSENSUS: [EQ]-x(4)-H-x(5)-[GSTA]-x(3)-[FY]-x(3)-[AG]-x(2)-[AV]-H-x(7)-P.

NAME: Photosystem I psaA and psaB proteins signature.
 CONSENSUS: C-D-G-P-G-R-G-G-T-C.

NAME: Photosystem I psaG and psaK proteins signature.
 CONSENSUS: G-F-x-[LIVM]-x-[DEA]-x(2)-[GA]-x-[GTA]-[SA]-x-G-H-x-[LIVM]-[GA].

NAME: Phytochrome chromophore attachment site signature.
 CONSENSUS: [RGS]-[GSA]-[PV]-H-x-C-H-x(2)-Y.

NAME: Phytochrome chromophore attachment site domain profile.

NAME: Speract receptor repeated domain signature.

CONSENSUS: G-x(5)-G-x(2)-E-x(6)-W-G-x(2)-C-x(3)-{FYW}-x(8)-C-x(3)-G.

NAME: TonB-dependent receptor proteins signature 1.

CONSENSUS: < x(10,115)-[DENF]-[ST]-[LIVMF]-[LIVSTEQ]-V-x-[AGP]-[STANEQPK].

NAME: TonB-dependent receptor proteins signature 2.

CONSENSUS: [LYGSTANE]-x(3)-[GSTAENQ]-x-[PGE]-R-x-[LIVFYWA]-x-[LIVMFTA]-[STAGNQ]-

CONSENSUS: [LIVMFYGT]-x-[LIVMFYWGTDQ]-x-F>.

NAME: Transmembrane 4 family signature.

CONSENSUS: G-x(3)-[LIVMF]-x(2)-[GSA]-[LIVMF](2)-G-C-x-[GA]-[STA]-x(2)-[EG]-x(2)-

CONSENSUS: [CWN]-[LIVM](2).

NAME: Bacterial chemotaxis sensory transducers signature.

CONSENSUS: R-T-E-[EQ]-Q-x(2)-[SA]-[LIVM]-x-[EQ]-T-A-A-S-M-E-Q-L-T-A-T-V.

NAME: ER lumen protein retaining receptor signature 1.

CONSENSUS: G-I-S-x-[KR]-x-Q-x-L-[FY]-x-[LIV](2)-F-x(2)-R-Y.

NAME: ER lumen protein retaining receptor signature 2.

CONSENSUS: L-E-[SA]-V-A-I-[LM]-P-Q-L.

NAME: Ephrins signature.

CONSENSUS: [KRQ]-[LF]-[CST]-x-K-[IF]-Q-x-[FY]-[ST]-[PA]-x(3)-G-x-E-F-x(5)-[FY](2)-

CONSENSUS: x(2)-[SA].

NAME: Granulins signature.

CONSENSUS: C-x-D-x(2)-H-C-C-P-x(4)-C.

NAME: HBGF/FGF family signature.

CONSENSUS: G-x-L-x-[STAGP]-x(6,7)-[DE]-C-x-[FM]-x-E-x(6)-Y.

NAME: PTN/MK heparin-binding protein family signature 1.

CONSENSUS: S-[DE]-C-x-[DE]-W-x-W-x(2)-C-x-P-x-[SN]-x-D-C-G-[LIVMA]-G-x-R-E-G.

NAME: PTN/MK heparin-binding protein family signature 2.

CONSENSUS: C-[KR]-[LIVM]-P-C-N-W-K-K-x-F-G-A-[DE]-C-K-Y-x-F-[EQ]-x-W-G-x-C.

NAME: Nerve growth factor family signature.

CONSENSUS: G-C-[KR]-G-[LIV]-[DE]-x(3)-[YW]-x-S-x-C.

NAME: Platelet-derived growth factor (PDGF) family signature.

CONSENSUS: P-[PS]-C-V-x(3)-R-C-[GSTA]-G-C-C.

NAME: Small cytokines (intercrine/chemokine) C-x-C subfamily signature.

CONSENSUS: C-x-C-[LIVM]-x(5,6)-[LIVMFY]-x(2)-[RKSEQ]-x-[LIVM]-x(2)-[LIVM]-x(5)-

CONSENSUS: [SAG]-x(2)-C-x(3)-[EQ]-[LIVM](2)-x(9,10)-C-L-[DN].

NAME: Small cytokines (intercrine/chemokine) C-C subfamily signature.

CONSENSUS: C-C-[LIFYT]-x(5,6)-[LI]-x(4)-[LIVMF]-x(2)-[FYW]-x(6,8)-C-x(3,4)-[SAG]-

CONSENSUS: [LIVM](2)-[FL]-x(8)-C-[STA].

NAME: TGF-beta family signature.

CONSENSUS: [LIVM]-x(2)-P-x(2)-[FY]-x(4)-C-x-G-x-C.

NAME: TNF family signature.

CONSENSUS: [LV]-x-[LIVM]-x(3)-G-[LIVMF]-Y-[LIVMFY](2)-x(2)-[QEKHL]-[LIVMGT]-x-

CONSENSUS: [LIVMFY].

NAME: TNF family profile.

NAME: Wnt-1 family signature.

CONSENSUS: C-K-C-H-G-[LIVMT]-S-G-x-C.

NAME: Interferon alpha, beta and delta family signature.

CONSENSUS: [FYH]-[FY]-x-[GNRC]-[LIVM]-x(2)-[FY]-L-x(7)-[CY]-A-W.

NAME: Granulocyte-macrophage colony-stimulating factor signature.

CONSENSUS: C-P-[LP]-T-x-E-[ST]-x-C.

NAME: Interleukin-1 signature.

CONSENSUS: [FC]-x-S-[ASLV]-x(2)-P-x(2)-[FYLV]-[LI]-[SCA]-T-x(7)-[LIVM].

NAME: Interleukin-2 signature.
 CONSENSUS: T-E-[LF]-x(2)-L-x-C-L-x(2)-E-L.

NAME: Interleukins -4 and -13 signature.
 CONSENSUS: L-x-E-[LIVM](2)-x(4,5)-[LIVM]-[TL]-x(5,7)-C-x(4)-[IVA]-x-[DNS]-[LIVMA].

NAME: Interleukin-6 / G-CSF / MGF signature.
 CONSENSUS: C-x(9)-C-x(6)-G-L-x(2)-[FY]-x(3)-L.

NAME: Interleukin-7 and -9 signature.
 CONSENSUS: N-x-[LAP]-[SCT]-F-L-K-x-L-L.

NAME: Interleukin-10 family signature.
 CONSENSUS: [GS]-C-x(2)-[LV]-x(2)-[LIVM](2)-x-F-Y-L-x(2)-V.

NAME: LIF / OSM family signature.
 CONSENSUS: [PST]-x(4)-F-[NQ]-x-K-x(3)-C-x-[LF]-L-x(2)-Y-[HK].

NAME: Macrophage migration inhibitory factor family signature.
 CONSENSUS: [DE]-P-C-A-x(3)-[LIVM]-x-S-I-G-x-[LIVM]-G.

NAME: Adipokinetin hormone family signature.
 CONSENSUS: Q-[LV]-[NT]-[FY]-[ST]-x(2)-W.

NAME: Bombesin-like peptides family signature.
 CONSENSUS: W-A-x-G-[SH]-[LF]-M.

NAME: Calcitonin / CGRP / IAPP family signature.
 CONSENSUS: C-[SAGDN]-[STN]-x(0,1)-[SA]-T-C-[VMA]-x(3)-[LYF]-x(3)-[LYF].

NAME: Corticotropin-releasing factor family signature.
 CONSENSUS: [PQ]-x-[LIVM]-S-[LIVM]-x(2)-[PST]-[LIVMF]-x-[LIVM]-L-R-x(2)-[LIVM].

NAME: Crustacean CHH/MIH/GIH neurohormones family signature.
 CONSENSUS: C-[DENK]-D-C-x-N-[LIV]-[FY]-R-x(7)-C-[KR]-x(2)-C.

NAME: Erythropoietin / thrombopoietin signature.
 CONSENSUS: P-x(4)-C-D-x-R-[LIVM](2)-x-[KR]-x(14)-C.

NAME: Granins signature 1.
 CONSENSUS: [DE]-[SN]-L-[SAN]-x(2)-[DE]-x-E-L.

NAME: Granins signature 2.
 CONSENSUS: C-[LIVM](2)-E-[LIVM](2)-S-[DN]-[STA]-L-x-K-x-S-x(3)-[LIVM]-[STA]-x-E-C.

NAME: Galanin signature.
 CONSENSUS: G-W-T-L-N-S-A-G-Y-L-L-G-P-H.

NAME: Gastrin / cholecystokinin family signature.
 CONSENSUS: Y-x(0,1)-[GD]-[WH]-M-[DR]-F.

NAME: Glucagon / GIP / secretin / VIP family signature.
 CONSENSUS: [YH]-[STAIVGD]-[DEQ]-[AGF]-[LIVMSTE]-[FYLR]-x-[DENSTAK]-[DENSTA]-
 CONSENSUS: [LIVMFYGG]-x(9)-[KREQL]-[KRDENQL]-[LVFYWG]-[LIVQ].

NAME: Glycoprotein hormones alpha chain signature 1.
 CONSENSUS: C-x-G-C-C-[FY]-S-R-A-[FY]-P-T-P.

NAME: Glycoprotein hormones alpha chain signature 2.
 CONSENSUS: N-H-T-x-C-x-C-x-T-C-x(2)-H-K.

NAME: Glycoprotein hormones beta chain signature 1.
 CONSENSUS: C-[STAGM]-G-[HFYL]-C-x-[ST].

NAME: Glycoprotein hormones beta chain signature 2.
 CONSENSUS: [PA]-V-A-x(2)-C-x-C-x(2)-C-x(4)-[STD]-[DEY]-C-x(6,8)-[PGSTAVM]-x(2)-C.

NAME: Gonadotropin-releasing hormones signature.
 CONSENSUS: Q-H-[FYW]-S-x(4)-P-G.

NAME: Insulin family signature.
 CONSENSUS: C-C-[P]-x(2)-C-[STDNEKPI]-x(3)-[LIVMFS]-x(3)-C.

NAME: Natriuretic peptides signature.
 CONSENSUS: C-F-G-x(3)-D-R-I-x(3)-S-x(2)-G-C.

NAME: Neurohypophysial hormones signature.
 CONSENSUS: C-[LIFY](2)-x-N-[CS]-P-x-G.

NAME: Neuromedin U signature.
 CONSENSUS: F-[LIVMF]-F-R-P-R-N.

NAME: Endogenous opioids neuropeptides precursors signature.
 CONSENSUS: C-x(3)-C-x(2)-C-x(2)-[KRH]-x(6,7)-[LIF]-[DN]-x(3)-C-x-[LIVM]-[EQ]-C.
 CONSENSUS: [EQ]-x(8)-W-x(2)-C.

NAME: Pancreatic hormone family signature.
 CONSENSUS: [FY]-x(3)-[LIVM]-x(2)-Y-x(3)-[LIVMFY]-x-R-x-R-[YF].

NAME: Parathyroid hormone family signature.
 CONSENSUS: V-S-E-x-Q-x(2)-H-x(2)-G.

NAME: Pyrokinins signature.
 CONSENSUS: F-[GSTV]-P-R-L-[G>].

NAME: Somatotropin, prolactin and related hormones signature 1.
 CONSENSUS: C-x-[ST]-x(2)-[LIVMFY]-x-[LIVMSTA]-P-x(5)-[TALIV]-x(7)-[LIVMFY]-x(6)-
 CONSENSUS: [LIVMFY]-x(2)-[STA]-W.

NAME: Somatotropin, prolactin and related hormones signature 2.
 CONSENSUS: C-[LIVMFY]-x(2)-D-[LIVMFYSTA]-x(5)-[LIVMFY]-x(2)-[LIVMFYT]-x(2)-C.

NAME: Tachykinin family signature.
 CONSENSUS: F-[IVFY]-G-[LM]-M-[G>].

NAME: Thymosin beta-4 family signature.
 CONSENSUS: K-L-K-K-T-E-T-Q-E-K-N.

NAME: Urotensin II signature.
 CONSENSUS: C-F-W-K-Y-C.

NAME: Cecropin family signature.
 CONSENSUS: W-x(0,2)-[KDN]-x(2)-K-[KRE]-[LI]-E-[RKN].

NAME: Mammalian defensins signature.
 CONSENSUS: C-x-C-x(3,5)-C-x(7)-G-x-C-x(9)-C-C.

NAME: Arthropod defensins signature.
 CONSENSUS: C-x(2,3)-[HN]-C-x(3,4)-[GR]-x(2)-G-G-x-C-x(4,7)-C-x-C.

NAME: Cathelicidins signature 1.
 CONSENSUS: Y-x-[ED]-x-V-x-[RQ]-A-[LIVMA]-[DQG]-x-[LIVMFY]-N-[EQ].

NAME: Cathelicidins signature 2.
 CONSENSUS: F-x-[LIVM]-K-E-T-x-C-x(10)-C-x-F-[KR]-[KE].

NAME: Endothelin family signature.
 CONSENSUS: C-x-C-x(4)-D-x(2)-C-x(2)-[FY]-C.

NAME: Plant thionins signature.
 CONSENSUS: C-C-x(5)-R-x(2)-[FY]-x(2)-C.

NAME: Gamma-thionins family signature.
 CONSENSUS: [KR]-x-C-x(3)-[SV]-x(2)-[FYWH]-x-[GF]-x-C-x(5)-C-x(3)-C.

NAME: Snake toxins signature.
 CONSENSUS: G-C-x(1,3)-C-P-x(8,10)-C-C-x(2)-[PDEN].

NAME: Myotoxins signature.
 CONSENSUS: K-x-C-H-x-K-x(2)-H-C-x(2)-K-x(3)-C-x(8)-K-x(2)-C-x(2)-[RK]-x-K-C-C-K-K.

NAME: Scorpion short toxins signature.
 CONSENSUS: C-x(3)-C-x(6,9)-[GAS]-K-C-[IMQT]-x(3)-C-x-C.

NAME: Heat-stable enterotoxins signature.
 CONSENSUS: C-C-x(2)-C-C-x-P-A-C-x-G-C.

NAME: Aerolysin type toxins signature.
 CONSENSUS: [KT]-x(2)-N-W-x(2)-T-[DN]-T.

NAME: Shiga/ricin ribosomal inactivating toxins active site signature.
 CONSENSUS: [LIVMA]-x-[LIVMSTA](2)-x-E-[SAGV]-[STAL]-R-[FY]-[RKNQS]-x-[LIVM]-[EQS]-
 CONSENSUS: x(2)-[LIVMF].

NAME: Channel forming colicins signature.
 CONSENSUS: T-x(2)-W-x-P-[LIVMFY](3)-x(2)-E.

NAME: Hok/gef family cell toxic proteins signature.
 CONSENSUS: [LIVMA](4)-C-[LIVMFA]-T-[LIVMA](2)-x(4)-[LIVM]-x-[RG]-x(2)-L-[CY].

NAME: Staphylococcal enterotoxin/Streptococcal pyrogenic exotoxin signature 1.
 CONSENSUS: Y-G-G-[LIV]-T-x(4)-N.

NAME: Staphylococcal enterotoxin/Streptococcal pyrogenic exotoxin signature 2.
 CONSENSUS: K-x(2)-[LIV]-x(4)-[LIV]-D-x(3)-R-x(2)-L-x(5)-[LIV]-Y.

NAME: Thiol-activated cytolysins signature.
 CONSENSUS: [RK]-E-C-T-G-L-x-W-E-W-W-[RK].

NAME: Membrane attack complex components / perforin signature.
 CONSENSUS: Y-x(6)-[FY]-G-T-H-[FY].

NAME: Pancreatic trypsin inhibitor (Kunitz) family signature.
 CONSENSUS: F-x(3)-G-C-x(6)-[FY]-x(5)-C.

NAME: Bowman-Birk serine protease inhibitors family signature.
 CONSENSUS: C-x(5,6)-[DENQKRHSTA]-C-[PASTDH]-[PASTDK]-[ASTDV]-C-[NDKS]-[DEKRHSTA]-C.

NAME: Kazal serine protease inhibitors family signature.
 CONSENSUS: C-x(7)-C-x(6)-Y-x(3)-C-x(2,3)-C.

NAME: Soybean trypsin inhibitor (Kunitz) protease inhibitors family signature.
 CONSENSUS: [LIVM]-x-D-x-[EDNTY]-[DG]-[RKHDENQ]-x-[LIVM]-x(5)-Y-x-[LIVM].

NAME: Serpins signature.
 CONSENSUS: [LIVMFY]-x-[LIVMFYAC]-[DNQ]-[RKHQS]-[PST]-F-[LIVMFY]-[LIVMFYC]-x-
 CONSENSUS: [LIVMFAH].

NAME: Potato inhibitor I family signature.
 CONSENSUS: [FYW]-P-[EQH]-[LIV](2)-G-x(2)-[STAGV]-x(2)-A.

NAME: Squash family of serine protease inhibitors signature.
 CONSENSUS: C-P-x(5)-C-x(2)-D-x-D-C-x(3)-C-x-C.

NAME: Streptomyces subtilisin-type inhibitors signature.
 CONSENSUS: C-x-P-x(2,3)-G-x-H-P-x(4)-A-C-[ATD]-x-L.

NAME: Cysteine proteases inhibitors signature.
 CONSENSUS: [GSTEQKRV]-Q-[LIVT]-[VAF]-[SAGQ]-G-x-[LIVMNK]-x(2)-[LIVMFY]-x-[LIVMFYA]-
 CONSENSUS: [DENQKRHSIV].

NAME: Tissue inhibitors of metalloproteinases signature.
 CONSENSUS: C-x-C-x-P-x-H-P-Q-x-A-F-C.

NAME: Cereal trypsin/alpha-amylase inhibitors family signature.
 CONSENSUS: C-x(4)-[SAGD]-x(4)-[SPAL]-[LF]-x(2)-C-[RH]-x-[LIVMFY](2)-x(3,4)-C.

NAME: Alpha-2-macroglobulin family thiolester region signature.
 CONSENSUS: [PG]-x-[GS]-C-[GA]-E-[EQ]-x-[LIVM].

NAME: Disintegrins signature.
 CONSENSUS: C-x(2)-G-x-C-C-x-[NQRS]-C-x-[FM]-x(6)-C-[RK].

NAME: Lambdoid phages regulatory protein CIII signature.
 CONSENSUS: E-S-x-L-x-R-x(2)-[KR]-x-L-x(4)-[KR](2)-x(2)-[DE]-x-L.

NAME: Chaperonins cpn60 signature.
 CONSENSUS: A-[AS]-x-[DEQ]-E-x(4)-G-G-[GA].

NAME: Chaperonins cpn10 signature.
 CONSENSUS: [LIVMFY]-x-P-[ILT]-x-[DEN]-[KR]-[LIVMFA](3)-[KREQ]-x(8,9)-[SG]-x-

CONSENSUS: [LIVMFY](3).

NAME: Chaperonins TCP-1 signature 1.

CONSENSUS: [RKEL]-[ST]-x-[LMFY]-G-P-x-[GSA]-x-x-K-[LIVMF](2).

NAME: Chaperonins TCP-1 signature 2.

CONSENSUS: [LIVM]-[TS]-[NK]-D-[GA]-[AVNHK]-[TAV]-[LIVM](2)-x(2)-[LIVM]-x-[LIVM]-x-

CONSENSUS: [SNH]-[PQH].

NAME: Chaperonins TCP-1 signature 3.

CONSENSUS: Q-[DEK]-x-x-[LIVMGTA]-[GA]-D-G-T.

NAME: Heat shock hsp20 proteins family profile.

NAME: Heat shock hsp70 proteins family signature 1.

CONSENSUS: [IV]-D-L-G-T-[ST]-x-[SC].

NAME: Heat shock hsp70 proteins family signature 2.

CONSENSUS: [LIVMF]-[LIVMFY]-[DN]-[LIVMFS]-G-[GSH]-[GS]-[AST]-x(3)-[ST]-[LIVM]-

CONSENSUS: [LIVMFC].

NAME: Heat shock hsp70 proteins family signature 3.

CONSENSUS: [LIVMY]-x-[LIVMF]-x-G-G-x-[ST]-x-[LIVM]-P-x-[LIVM]-x-[DEQKRSTA].

NAME: Heat shock hsp90 proteins family signature.

CONSENSUS: Y-x-[NQH]-K-[DE]-[IVA]-F-L-R-[ED].

NAME: Chaperonins clpA/B signature 1.

CONSENSUS: D-[AI]-[SGA]-N-[LIVMF](2)-K-[PT]-x-L-x(2)-G.

NAME: Chaperonins clpA/B signature 2.

CONSENSUS: R-[LIVMFY]-D-x-S-E-[LIVMFY]-x-E-[KRQ]-x-[STA]-x-[STA]-[KR]-[LIVM]-x-G-

CONSENSUS: [STA].

NAME: Nt-dnaJ domain signature.

CONSENSUS: [FY]-x(2)-[LIVMA]-x(3)-[FYWHNT]-[DENQSA]-x-L-x-[DN]-x(3)-[KR]-x(2)-[FYI].

NAME: dnaJ domain profile.

NAME: CXXCXGXXG dnaJ domain signature.

CONSENSUS: C-[DEGSTHKR]-x-C-x-G-x-[GK]-[AGSDM]-x(2)-[GSNKR]-x(4,6)-C-x(2,3)-C-x-G-x-G.

NAME: grpE protein signature.

CONSENSUS: [FL]-[DN]-[PHEA]-x(2)-[HM]-x-A-[LIVMTN]-x(16,20)-G-[FY]-x(3)-[DEG]-x(2)-

CONSENSUS: [LIVM]-[RI]-x-[SA]-x-V-x-[IV].

NAME: Bacterial type II secretion system protein C signature.

CONSENSUS: P-x(6)-F-x(4)-L-x(3)-D-[LIVM]-A-[LIVM]-x-[LIVM]-N-x-[LIVM]-x-L.

NAME: Bacterial type II secretion system protein D signature.

CONSENSUS: [GR]-[DEQKG]-[STVM]-[LIVMA](3)-[GA]-G-[LIVMFY]-x(11)-[LIVM]-P-

CONSENSUS: [LIVMFYWGS]-[LIVMF]-[GSAE]-x-[LIVM]-P-[LIVMFYW](2)-x(2)-[LV]-F.

NAME: Bacterial type II secretion system protein E signature.

CONSENSUS: [LIVM]-R-x(2)-P-D-x-[LIVM](3)-G-E-[LIVM]-R-D.

NAME: Bacterial type II secretion system protein F signature.

CONSENSUS: [KRQ]-[LIVMA]-x(2)-[SAIV]-[LIVM]-x-[TY]-P-x(2)-[LIVM]-x(3)-[STAGV]-x(6)-

CONSENSUS: [LMY]-x(3)-[LIVMF](2)-P.

NAME: Bacterial type II secretion system protein N signature.

CONSENSUS: G-T-L-W-x-G-x(11)-L-x(4)-W.

NAME: Bacterial export FHIPEP family signature.

CONSENSUS: R-[LIVM]-[GSA]-E-V-[GSA]-A-R-F-[STV]-L-D-[GSA]-M-P-G-K-Q-M-[GSA]-I-D-

CONSENSUS: [GSA]-D.

NAME: Protein secA signatures.

CONSENSUS: [IV]-x-[IV]-[SA]-T-[NQ]-M-A-G-R-G-x-D-I-x-L.

NAME: Protein secY signature 1.

CONSENSUS: [GST]-[LIVMF](2)-x-[LIVM]-G-[LIVM]-x-P-[LIVMFY](2)-x-[AS]-[GSTQ]-

CONSENSUS: [LIVMFAT](3)-Q-[LIVMFA](2).

NAME: Protein secY signature 2.
 CONSENSUS: [LIVMFYW](2)-x-[DE]-x-[LIVMF]-[STN]-x(2)-G-[LIVMF]-[GST]-[NST]-G-x-[GST]-
 CONSENSUS: [LIVMF](3).

NAME: Protein secE/sec61-gamma signature.
 CONSENSUS: [LIVMFY]-x(2)-[DENQGA]-x(4)-[LIVMTA]-x-[KRV]-x(2)-[KW]-P-x(3)-[SEQ]-x(7)-
 CONSENSUS: [LIVT]-[LIVGA]-[LIVFGAST].

NAME: Gram-negative pili assembly chaperone signature.
 CONSENSUS: [LIVMFY]-[APN]-x-[DNS]-[KREQ]-E-[STR]-[LIVMAR]-x-[FYWT]-x-[NC]-[LIVM]-
 CONSENSUS: x(2)-[LIVM]-P-[PAS].

NAME: Fimbrial biogenesis outer membrane usher protein signature.
 CONSENSUS: [VL]-[PASQ]-[PAS]-G-[PAD]-[FY]-x-[LI]-[DNQSTAP]-[DNH]-[LIVMFY].

NAME: SRP54-type proteins GTP-binding domain signature.
 CONSENSUS: P-[LIVM]-x-[FYL]-[LIVMAT]-[GS]-x-[GS]-[EQ]-x(4)-[LIVMF].

NAME: Cytochrome c oxidase assembly factor COX10/ctaB/cyoE signature.
 CONSENSUS: [ED]-x-D-x(2)-M-x-R-T-x(2)-R-x(4)-G.

NAME: Cyclin-dependent kinases regulatory subunits signature 1.
 CONSENSUS: Y-S-x-[KR]-Y-x-[DE](2)-x-[FY]-E-Y-R-H-V-x-[LV]-[PT]-[KRP].

NAME: Cyclin-dependent kinases regulatory subunits signature 2.
 CONSENSUS: H-x-P-E-x-H-[IV]-L-L-F-[KR].

NAME: Pentaxin family signature.
 CONSENSUS: H-x-C-x-[ST]-W-x-[ST].

NAME: Immunoglobulins and major histocompatibility complex proteins signature.
 CONSENSUS: [FY]-x-C-x-[VA]-x-H.

NAME: Prion protein signature 1.
 CONSENSUS: A-G-A-A-A-G-A-V-V-G-G-L-G-G-Y.

NAME: Prion protein signature 2.
 CONSENSUS: E-x-[ED]-x-K-[LIVM](2)-x-[KR]-[LIVM](2)-x-[QE]-M-C-x(2)-Q-Y.

NAME: Cyclins signature.
 CONSENSUS: R-x(2)-[LIVMSA]-x(2)-[FYWS]-[LIVM]-x(8)-[LIVMFC]-x(4)-[LIVMFYA]-x(2)-
 CONSENSUS: [STAGC]-[LIVMFYQ]-x-[LIVMFYC]-[LIVMFY]-D-[RKH]-[LIVMFYW].

NAME: Proliferating cell nuclear antigen signature 1.
 CONSENSUS: [GA]-[LIVMF]-x-[LIVMA]-x-[SAV]-[LIVM]-D-x-[NSAE]-[HKR]-[VI]-x-[LY]-
 CONSENSUS: [VGA]-x-[LIVM]-x-[LIVM]-x(4)-F.

NAME: Proliferating cell nuclear antigen signature 2.
 CONSENSUS: [RKA]-C-[DE]-[RH]-x(3)-[LIVMF]-x(3)-[LIVM]-x-[SGAN]-[LIVMF]-x-K-
 CONSENSUS: [LIVMF](2).

NAME: Actin-depolymerizing proteins signature.
 CONSENSUS: P-[DE]-x-[SA]-x-[LIVMT]-[KR]-x-[KR]-M-[LIVM]-[YA]-[STA](3)-x(3)-[LIVMF]-
 CONSENSUS: [KR].

NAME: BCL2-like apoptosis inhibitors (spans part of BH3, BH1 and BH2).

NAME: Apoptosis regulator, Bcl-2 family BH1 domain signature.
 CONSENSUS: [LVME]-[FT]-x-[GSD]-[GL]-x(1,2)-[NS]-[YW]-G-R-[LIV]-[LIVC]-[GAT]-
 CONSENSUS: [LIVMF](2)-x-F-[GSAE]-[GSARY].

NAME: Apoptosis regulator, Bcl-2 family BH2 domain signature.
 CONSENSUS: W-[LIM]-x(3)-[GR]-G-[WQ]-[DENSAY]-x-[FLGA]-[LIVFTC].

NAME: Apoptosis regulator, Bcl-2 family BH3 domain signature.
 CONSENSUS: [LIVAT]-x(3)-L-[KARQ]-x-[IVAL]-G-D-[DESG]-[LIMFV]-[DENS HQ]-[LVSHRQ]-
 CONSENSUS: [NSR].

NAME: Apoptosis regulator, Bcl-2 family BH4 domain signature.
 CONSENSUS: [DS]-[NT]-R-[AE]-[LI]-V-x-[KD]-[FY]-[LIV]-[GHS]-Y-K-L-[SR]-Q-[RK]-G-
 CONSENSUS: [HY]-x-[CW].

NAME: Apoptosis regulator, Bcl-2 family BH4 domain profile.

NAME: Arrestins signature.

CONSENSUS: [FY]-R-Y-G-x-[DE](2)-x-[DE]-[LIVM](2)-G-[LIVM]-x-F-x-[RK]-[DEQ]-[LIVM].

NAME: AAA-protein family signature.

CONSENSUS: [LIVMT]-x-[LIVMT]-[LIVMF]-x-[GATMC]-[ST]-[NS]-x(4)-[LIVM]-D-x-A-[LIFA]-

CONSENSUS: x-R.

NAME: Ubiquitin domain signature.

CONSENSUS: K-x(2)-[LIVM]-x-[DESAK]-x(3)-[LIVM]-[PA]-x(3)-Q-x-[LIVM]-[LIVMC]-

CONSENSUS: [LIVMFY]-x-G-x(4)-[DE].

NAME: Ubiquitin domain profile.

NAME: ADP-ribosylation factors family signature.

CONSENSUS: [HRQT]-x-[FYWI]-x-[LIVM]-x(4)-A-x(2)-G-x(2)-[LIVM]-x(2)-[GSA]-[LIVMF]-x-

CONSENSUS: [WK]-[LIVM].

NAME: GTP-binding nuclear protein ran signature.

CONSENSUS: D-T-A-G-Q-E-K-[LF]-G-G-L-R-[DE]-G-Y-Y.

NAME: SAR1 family signature.

CONSENSUS: R-x-[LIVM]-E-V-F-M-C-S-[LIVM](2)-x-[KRQ]-x-G-Y-x-E-[AG]-[FT]-x-W-[LIVM]-

CONSENSUS: x-Q-Y.

NAME: Band 7 protein family signature.

CONSENSUS: R-x(2)-[LIV]-[SAN]-x(6)-[LIV]-D-x(2)-T-x(2)-W-G-[LIV]-[KRH]-[LIV]-x-

CONSENSUS: [KR]-[LIV]-E-[LIV]-[KR].

NAME: Trp-Asp (WD) repeats signature.

CONSENSUS: [LIVMSTAC]-[LIVMFYWSTAGC]-[LIMSTAG]-[LIVMSTAGC]-x(2)-[DN]-x(2)-

CONSENSUS: [LIVMWTAC]-x-[LIVMFSTAG]-W-[DEN]-[LIVMFSTAGCN].

NAME: G-protein gamma subunit profile.

NAME: Ras GTPase-activating proteins signature.

CONSENSUS: [GSN]-x-[LIVMF]-[FY]-[LIVMFY]-R-[LIVMFY](2)-[GACN]-P-[AV]-[LIV](2)-

CONSENSUS: [SGAN]-P.

NAME: Ras GTPase-activating proteins profile.

NAME: Guanine-nucleotide dissociation stimulators CDC24 family signature.

CONSENSUS: L-x(2)-[LIVMFYW]-L-x(2)-P-[LIVM]-x(2)-[LIVM]-x-[KRS]-x(2)-L-x-[LIVM]-x-

CONSENSUS: [DEQ]-[LIVM]-x(3)-[ST].

NAME: Guanine-nucleotide dissociation stimulators CDC25 family signature.

CONSENSUS: [GAP]-[CT]-V-P-[FY]-x(4)-[LIVMFY]-x-[DN]-[LIVM].

NAME: MARCKS family signature 1.

CONSENSUS: G-Q-E-N-G-H-V-[KR].

NAME: MARCKS family phosphorylation site domain.

CONSENSUS: E-T-P-K(5)-x(0,1)-F-S-F-K-K-x-F-K-L-S-G-x-S-F-K-[KR]-[NS]-[KR]-K-E.

NAME: Stathmin family signature 1.

CONSENSUS: P-[KQ]-[KR](2)-[DE]-x-S-L-[EG]-E.

NAME: Stathmin family signature 2.

CONSENSUS: A-E-K-R-E-H-E-[KR]-E-V.

NAME: GTP-binding elongation factors signature.

CONSENSUS: D-[KRSTGANQFYW]-x(3)-E-[KRAQ]-x-[RKQD]-[GC]-[IVMK]-[ST]-[IV]-x(2)-

CONSENSUS: [GSTACKRNQ].

NAME: Elongation factor 1 beta/beta'/delta chain signature 1.

CONSENSUS: [DE]-[DEG]-[DE](2)-[LIVMF]-D-L-F-G.

NAME: Elongation factor 1 beta/beta'/delta chain signature 2.

CONSENSUS: V-Q-S-x-D-[LIVM]-x-A-[FWM]-[NQ]-K-[LIVM].

NAME: Elongation factor 1 gamma chain profile.

NAME: Elongation factor Ts signature 1.

CONSENSUS: L-R-x(2)-T-[GDQ]-x-[GS]-[LIVMF]-x(0,1)-[DENKAC]-x-K-[KRNEQS]-[AV]-L.

NAME: Elongation factor Ts signature 2.
 CONSENSUS: E-[LIVM]-N-[SCV]-[QE]-T-D-F-V-[SA]-[KRN].

NAME: Elongation factor P signature.
 CONSENSUS: K-x-A-x(4)-G-x(2)-[LIV]-x-V-P-x(2)-[LIV]-x(2)-G.

NAME: Eukaryotic initiation factor 1A signature.
 CONSENSUS: [IM]-x-G-x-[GS]-[KRH]-x(4)-[CL]-x-D-G-x(2)-R-x(2)-[RH]-I-x-G.

NAME: Eukaryotic initiation factor 4E signature.
 CONSENSUS: [DE]-[IFY]-x(2)-F-[KR]-x(2)-[LIVM]-x-P-x-W-E-[DV]-x(5)-G-G-[KR]-W.

NAME: Eukaryotic initiation factor 5A hypusine signature.
 CONSENSUS: [PT]-G-K-H-G-x-A-K.

NAME: Initiation factor 2 signature.
 CONSENSUS: G-x-[LIVM]-x(2)-L-[KR]-[KRHNS]-x-K-x(5)-[LIVM]-x(2)-G-x-[DEN]-C-G.

NAME: Initiation factor 3 signature.
 CONSENSUS: [KR]-[LIVM](2)-[DN]-[FY]-[GSN]-[KR]-[LIVMFYS]-x-[FY]-[DEQT]-x(2)-[KR].

NAME: Translation initiation factor SUI1 signature.
 CONSENSUS: [LIVM]-[EQ]-[LIVM]-Q-G-[DEN]-[KHQ]-[KRV].

NAME: Prokaryotic-type class I peptide chain release factors signature.
 CONSENSUS: [AR]-[STA]-x-G-x-G-G-Q-[HNGCS]-V-N-x(3)-[ST]-A-[IV].

NAME: Transcription termination factor nusG signature.
 CONSENSUS: [LIVM]-F-G-[KRW]-x-T-P-[IV]-x-[LIVM].

NAME: Calponin family repeat.
 CONSENSUS: [LIVM]-x-[LS]-Q-[MAS]-G-[STY]-[NT]-[KRQ]-x(2)-[STN]-Q-x-G-x(3,4)-G.

NAME: CAP protein signature 1.
 CONSENSUS: [LIVM](2)-x-R-L-[DE]-x(4)-R-L-E.

NAME: CAP protein signature 2.
 CONSENSUS: D-[LIVMFY]-x-E-x-[PA]-x-P-E-Q-[LIVMFY]-K.

NAME: Calreticulin family signature 1.
 CONSENSUS: [KRHN]-x-[DEQN]-[DEQNK]-x(3)-C-G-G-[AG]-[FY]-[LIVM]-[KN]-[LIVMFY](2).

NAME: Calreticulin family signature 2.
 CONSENSUS: [LIVM](2)-F-G-P-D-x-C-[AG].

NAME: Calreticulin family repeated motif signature.
 CONSENSUS: [IV]-x-D-x-[DENST]-x(2)-K-P-[DEH]-D-W-[DEN].

NAME: Calsequestrin signature 1.
 CONSENSUS: [EQ]-[DE]-G-L-[DN]-F-P-x-Y-D-G-x-D-R-V.

NAME: Calsequestrin signature 2.
 CONSENSUS: [DE]-L-E-D-W-[LIVM]-E-D-V-L-x-G-x-[LIVM]-N-T-E-D-D-D.

NAME: S-100/ICaBP type calcium binding protein signature.
 CONSENSUS: [LIVMFYW](2)-x(2)-[LK]-D-x(3)-[DN]-x(3)-[DNSG]-[FY]-x-[ES]-[FYVC]-x(2)-
 CONSENSUS: [LIVMFS]-[LIVMF].

NAME: Hemolysin-type calcium-binding region signature.
 CONSENSUS: D-x-[LI]-x(4)-G-x-D-x-[LI]-x-G-G-x(3)-D.

NAME: HlyD family secretion proteins signature.
 CONSENSUS: [LIVM]-x(2)-G-[LM]-x(3)-[STGAV]-x-[LIVMT]-x-[LIVMT]-[GE]-x-[KR]-x-
 CONSENSUS: [LIVMFYW](2)-x-[LIVMFYW](3).

NAME: P-II protein urydylatation site.
 CONSENSUS: Y-[KR]-G-[AS]-[AE]-Y.

NAME: P-II protein C-terminal region signature.
 CONSENSUS: [ST]-x(3)-G-[DY]-G-[KR]-[IV]-[FW]-[LIVM]-x(2)-[LIVM].

NAME: 14-3-3 proteins signature 1.
 CONSENSUS: R-N-L-[LIV]-S-[VG]-[GA]-Y-[KN]-N-[IVA].

NAME: 14-3-3 proteins signature 2.
 CONSENSUS: Y-K-[DE]-S-T-L-I-[IM]-Q-L-[LF]-[RHC]-D-N-[LF]-T-[LS]-W-[TAN]-[SAD].

NAME: ATP1G1 / PLM / MAT8 family signature.
 CONSENSUS: [DNS]-x-F-x-Y-D-x(2)-[ST]-[LIVM]-[RQ]-x(2)-G.

NAME: BTG1 family signature 1.
 CONSENSUS: Y-x(2)-[HP]-W-[FY]-[AP]-E-x-P-x-K-G-x-[GA]-[FY]-R-C-[IV]-[RH]-[IV].

NAME: BTG1 family signature 2.
 CONSENSUS: [LV]-P-x-[DE]-[LM]-[ST]-[LIVM]-W-[IV]-D-P-x-E-V-[SC]-x-[RQ]-x-G-E.

NAME: Cullin family signature.
 CONSENSUS: [LIV]-K-x(2)-[LIV]-x(2)-L-I-[DEQ]-[KRHNQ]-x-Y-[LIVM]-x-R-x(6,7)-[FY]-x-
 CONSENSUS: Y-x-[SA]>.

NAME: Cullin family profile.

NAME: Enhancer of rudimentary signature.
 CONSENSUS: Y-D-I-[SA]-x-L-[FY]-x-F-[IV]-D-x(3)-D-[LIV]-S.

NAME: G10 protein signature 1.
 CONSENSUS: L-C-C-x-[KR]-C-x(4)-[DE]-x-N-x(4)-C-x-C-R-V-P.

NAME: G10 protein signature 2.
 CONSENSUS: C-x-H-C-G-C-[KRH]-G-C-[SA].

NAME: Glucokinase regulatory protein family signature.
 CONSENSUS: G-[PA]-E-x-[LIV]-[STA]-G-S-[ST]-R-[LIVM]-K-[STGA](3)-x(2)-K.

NAME: GTP1/OBG family signature.
 CONSENSUS: D-[LIVM]-P-G-[LIVM](2)-[DEY]-[GN]-A-x(2)-G-x-G.

NAME: HIT family signature.
 CONSENSUS: [NQA]-x(4)-[GAV]-x-[QF]-x-[LIVM]-x-H-[LIVMFYT]-H-[LIVMFT]-H-[LIVMF](2)-
 CONSENSUS: [PSGA].

NAME: Caseins alpha/beta signature.
 CONSENSUS: C-L-[LV]-A-x-A-[LVF]-A.

NAME: Clathrin adaptor complexes medium chain signature 1.
 CONSENSUS: [IVT]-[GSP]-W-R-x(2,3)-[GAD]-x(2)-[HY]-x(2)-N-x-[LIVMAFY](3)-D-[LIVM]-
 CONSENSUS: [LIVMT]-E.

NAME: Clathrin adaptor complexes medium chain signature 2.
 CONSENSUS: [LIV]-x-F-I-P-P-x-G-x-[LIVMFY]-x-L-x(2)-Y.

NAME: Clathrin adaptor complexes small chain signature.
 CONSENSUS: [LIVM](2)-Y-[KR]-x(4)-L-Y-F.

NAME: Ependymins signature 1.
 CONSENSUS: F-E-E-G-x-[LIVMF]-Y-[ED]-I-D-x(2)-N-[QE]-S-C-[RKH](2).

NAME: Ependymins signature 2.
 CONSENSUS: [QE]-[LIVMA]-F-x(2)-P-[STA]-[FY]-C-[DE]-[GA]-[LIVM]-x(2)-[DE](2).

NAME: Syntaxin / epimorphin family signature.
 CONSENSUS: [RQ]-x(3)-[LIVMA]-x(2)-[LIVM]-[ESH]-x(2)-[LIVMT]-x-[DEV]-[LIVM]-x(2)-
 CONSENSUS: [LIVM]-[FS]-x(2)-[LIVM]-x(3)-[LIVT]-x(2)-Q-[GADEQ]-x(2)-[LIVM]-[DNQT]-x-
 CONSENSUS: [LIVMF]-[DESV]-x(2)-[LIVM].

NAME: Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 signature 1.
 CONSENSUS: [GDER]-H-[FYWH]-T-Q-[LIVM](2)-W-x(2)-[STN].

NAME: Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 signature 2.
 CONSENSUS: [LIVMFYH]-[LIVMFY]-x-C-[NQRHS]-Y-x-[PARH]-x-[GL]-N-[LIVMFYWDN].

NAME: Fetuin family signature 1.
 CONSENSUS: C-x(56)-C-x(10)-C-x(13)-C-x(17,18)-C-x(13)-C-x(2)-C-x(58)-C-x(10,11)-
 CONSENSUS: C-x(10,12)-C-x(16,22)-C.

NAME: Fetuin family signature 2.
 CONSENSUS: L-E-T-x-C-H-x-L-D-P-T-P.

NAME: Legume lectins beta-chain signature.
 CONSENSUS: [LIV]-[STAG]-V-[DEQV]-[FLI]-D-[ST].

NAME: Legume lectins alpha-chain signature.
 CONSENSUS: [LIV]-x-[EDQ]-[FYWKR]-V-x-[LIV]-G-[LF]-[ST].

NAME: Vertebrate galactoside-binding lectin signature.
 CONSENSUS: W-[GEK]-x-[EQ]-x-[KRE]-x(3,6)-[PCTF]-[LIVMF]-[NQEKGSKV]-x-[GH]-x(3)-
 CONSENSUS: [DENKHS]-[LIVMFC].

NAME: Lysosome-associated membrane glycoproteins duplicated domain signature.
 CONSENSUS: [STA]-C-[LIVM]-[LIVMFYW]-A-x-[LIVMFYW]-x(3)-[LIVMFYW]-x(3)-Y.

NAME: LAMP glycoproteins transmembrane and cytoplasmic domain signature.
 CONSENSUS: C-x(2)-D-x(3,4)-[LIVM](2)-P-[LIVM]-x-[LIVM]-G-x(2)-[LIVM]-x-G-[LIVM](2)-
 CONSENSUS: x-[LIVM](4)-A-[FY]-x-[LIVM]-x(2)-[KR]-[RH]-x(1,2)-[STAG](2)-Y-[EQ].

NAME: Glycophorin A signature.
 CONSENSUS: I-I-x-[GAC]-V-M-A-G-[LIVM](2).

NAME: PMP-22 / EMP / MP20 family signature 1.
 CONSENSUS: [LIVMF](4)-[SA]-T-x(2)-[DNKS]-x-W-x(9,13)-[LIV]-W-x(2)-C.

NAME: PMP-22 / EMP / MP20 family signature 2.
 CONSENSUS: [RQ]-[AV]-x-M-[IV]-L-S-x-[LI]-x(4)-[GSA]-[LIVMF](3).

NAME: Oxysterol-binding protein family signature.
 CONSENSUS: E-[KQ]-x-S-H-[HR]-P-P-x-[STACF]-A.

NAME: Yeast PIR proteins repeats signature.
 CONSENSUS: S-Q-[IV]-[STGNH]-D-G-Q-[LIV]-Q-[AIV]-[STA].

NAME: Seminal vesicle protein I repeats signature.
 CONSENSUS: [IVM]-x-G-Q-D-x-V-K-x(5)-[KN]-G-x(3)-[STLV].

NAME: Seminal vesicle protein II repeats signature.
 CONSENSUS: [GSA]-Q-x-K-S-[FY]-x-Q-x-K-[SA].

NAME: Serum amyloid A proteins signature.
 CONSENSUS: A-R-G-N-Y-[ED]-A-x-[QKR]-R-G-x-G-G-x-W-A.

NAME: Spermadhesins family signature 1.
 CONSENSUS: C-G-x(2)-[LI]-x(4)-G-x-I-x(9)-C-x-W-T.

NAME: Spermadhesins family signature 2.
 CONSENSUS: C-x-K-E-x-[LIVM]-E-[LIVM]-x-[DE]-x(3)-[GS]-x(5)-K-x-C.

NAME: Stress-induced proteins SRP1/TIP1 family signature.
 CONSENSUS: P-W-Y-[ST](2)-R-L.

NAME: Glypicans signature.
 CONSENSUS: C-x(2)-C-x-G-[LIVM]-x(4)-P-C-x(2)-[FY]-C-x(2)-[LIVM]-x(2)-G-C.

NAME: Syndecans signature.
 CONSENSUS: [FY]-R-[IM]-[KR]-K(2)-D-E-G-S-Y.

NAME: Tissue factor signature.
 CONSENSUS: W-K-x-K-C-x(2)-T-x-[DEN]-T-E-C-D-[LIVM]-T-D-E.

NAME: Translationally controlled tumor protein signature 1.
 CONSENSUS: [LA]-G-[GAS]-N-[PA]-S-A-E-[GDE]-[PAGE]-x(0,1)-[DEG]-x-[DEN]-x(2)-[DE].

NAME: Translationally controlled tumor protein signature 2.
 CONSENSUS: [FL]-[FY]-[IVT]-G-E-x-[MA]-x(2,5)-[DEN]-[GAS]-x-[LV]-[AV]-x(3)-[FY]-[KR]-
 CONSENSUS: [DE].

NAME: Tub family signature 1.
 CONSENSUS: F-[KHQ]-G-R-V-[ST]-x-A-S-V-K-N-F-Q.

NAME: Tub family signature 2.
 CONSENSUS: A-F-[AG]-I-[SAC]-[LIVM]-[ST]-S-F-x-[GST]-K-x-A-C-E.

NAME: HCP repeats signature.
 CONSENSUS: H-R-H-R-G-H-x(2)-[DE](7).

NAME: Bacterial ice-nucleation proteins octamer repeat.
 CONSENSUS: A-G-Y-G-S-T-x-T.

NAME: Cell cycle proteins fisW / rodA / spoVE signature.
 CONSENSUS: [NV]-x(5)-[GTR]-[LIVMA]-x-P-[PTLIVM]-x-G-[LIVM]-x(3)-[LIVMFW](2)-S-[YSA]-
 CONSENSUS: G-G-[STN]-[SA].

NAME: Enterobacterial virulence outer membrane protein signature 1.
 CONSENSUS: G-[LIVMFY]-N-[LIVM]-K-Y-R-Y-E.

NAME: Enterobacterial virulence outer membrane protein signature 2.
 CONSENSUS: [FYW]-x(2)-G-x-G-Y-[KR]-F>.

NAME: Hydrogenases expression/synthesis hypA family signature.
 CONSENSUS: F-[CSA]-[FY]-[DE]-[LIVA](2)-x(3)-[ST]-[LIVM]-x(16)-C-x(2)-C-x(12,15)-
 CONSENSUS: C-P-x-C.

NAME: Hydrogenases expression/synthesis hupF/hupC family signature.
 CONSENSUS: <M-C-[LIV]-[GA]-[LIV]-P-x-[QKR]-[LIV].

NAME: Staphylocoagulase repeat signature.
 CONSENSUS: A-R-P-x(3)-K-x-S-x-T-N-A-Y-N-V-T-T-x(2)-[DN]-G-x(3)-Y-G.

NAME: 11-S plant seed storage proteins signature.
 CONSENSUS: N-G-x-[DE](2)-x-[LIVMF]-C-[ST]-x(11,12)-[PAG]-D.

NAME: Dehydrins signature 1.
 CONSENSUS: S(5)-[DE]-x-[DE]-G-x(1,2)-G-x(0,1)-[KR](4).

NAME: Dehydrins signature 2.
 CONSENSUS: [KR]-[LIM]-K-[DE]-K-[LIM]-P-G.

NAME: Germin family signature.
 CONSENSUS: G-x(4)-H-x-H-P-x-A-x-E-[LIVM].

NAME: Oleosins signature.
 CONSENSUS: [AG]-[ST]-x(2)-[AG]-x(2)-[LIVM]-[SAD]-T-P-[LIVMF](4)-F-S-P-[LIVM](3)-
 CONSENSUS: P-A.

NAME: Small hydrophilic plant seed proteins signature.
 CONSENSUS: G-[EQ]-T-V-V-P-G-G-T.

NAME: Pathogenesis-related proteins BectI family signature.
 CONSENSUS: G-x(2)-[LIVMF]-x(4)-E-x(2)-[CSTAEN]-x(8,9)-[GND]-G-[GS]-[CS]-x(2)-K-x(4)-
 CONSENSUS: [FY].

NAME: Pollen proteins Ole e I family signature.
 CONSENSUS: [EQ]-G-x-V-Y-C-D-T-C-R.

NAME: Thaumatin family signature.
 CONSENSUS: G-x-[GF]-x-C-x-T-[GA]-D-C-x(1,2)-G-x(2,3)-C.

NAME: Mrp family signature.
 CONSENSUS: W-x(2)-[LIVM]-D-[LIVMY](4)-D-x-P-P-G-T-[GS]-D.

NAME: Glucose inhibited division protein A family signature 1.
 CONSENSUS: [GS]-P-x-Y-C-P-S-[LIVM]-E-x-K-[LIVM]-x-[KR]-F.

NAME: Glucose inhibited division protein A family signature 2.
 CONSENSUS: A-G-Q-x-[NT]-G-x(2)-G-Y-x-E-[SAG](3)-[QS]-G-[LIVM](2)-A-G-[LIVMT]-N-A.

NAME: NOL1/NOP2/sun family signature.
 CONSENSUS: [FV]-D-[KRA]-[LIVMA]-L-x-D-[AV]-P-C-[ST]-[GA].

NAME: PET112 family signature.
 CONSENSUS: [DN]-x-[DN]-R-x(3)-P-L-[LIV]-E-[LIV]-x-[ST]-x-P.

NAME: Protein smpB signature.
 CONSENSUS: [TA]-G-[LIVM]-x-L-x-G-x-E-[LIVM]-[KQ]-[SA]-[LIVM].

NAME: Hypothetical cof family signature 1.
 CONSENSUS: [LIVFYAN]-[LIVMFA]-x(2)-D-[LIVMF]-[ND]-G-T-[LIV]-[LVY]-[STANLM].

NAME: Hypothetical cof family signature 2.
 CONSENSUS: [LIVMFC]-G-D-[GSANQ]-x-N-D-x(3)-[LIMFY]-x(2)-[AV]-x(2)-[GSCP]-x(2)-
 CONSENSUS: [LMP]-x(2)-[GAS].

NAME: RIO1/ZK632.3/MJ0444 family signature.
 CONSENSUS: [LIVM]-V-H-[GA]-D-L-S-E-[FY]-N-x-[LIVM].

NAME: SUAS/yCiO/yrdC family signature.
 CONSENSUS: [LIVMTA](3)-[LIVMFYC]-[PG]-T-[DE]-[STA]-x-[FY]-[GA]-[LIVM]-[GS].

NAME: Uncharacterized protein family UPF0001 signature.
 CONSENSUS: [FW]-H-[FM]-[IV]-G-x-[LIV]-Q-x-[NKR]-K-x(3)-[LIV].

NAME: Uncharacterized protein family UPF0003 signature.
 CONSENSUS: G-x-V-x(2)-[LIV]-x(3)-[SA]-x(6)-D-x(3)-[LIVT](3)-P-N-x(2)-[LIVMF](2)-
 CONSENSUS: x(5)-N.

NAME: Uncharacterized protein family UPF0004 signature.
 CONSENSUS: [LIVM]-x-[LIVMT]-x(2)-G-C-x(3)-C-[STAN]-[FY]-C-x-[LIVM]-x(4)-G.

NAME: Uncharacterized protein family UPF0005 signature.
 CONSENSUS: G-[LIVM](2)-[SA]-x(5,8)-G-x(2)-[LIVM]-G-P-x-L-x(4)-[SAG]-x(4,6)-
 CONSENSUS: [LIVM](2)-x(2)-A-x(3)-T-A-[LIVM](2)-F.

NAME: Uncharacterized protein family UPF0006 signature 1.
 CONSENSUS: [LIVMFY](2)-D-[STA]-H-x-H-[LIVMF]-[DN].

NAME: Uncharacterized protein family UPF0006 signature 2.
 CONSENSUS: P-[LIVM]-x-[LIVM]-H-x-R-x-[TA]-x-[DE].

NAME: Uncharacterized protein family UPF0006 signature 3.
 CONSENSUS: [LVSA]-[LIVA]-x(2)-[LIVM]-[PS]-x(3)-L-[LIVM]-[LIVMS]-E-T-D-x-P.

NAME: Uncharacterized protein family UPF0007 signature.
 CONSENSUS: V-L-[IV]-H-D-[GA]-A-R.

NAME: Uncharacterized protein family UPF0011 signature.
 CONSENSUS: S-D-A-G-x-P-x-[LIV]-[SN]-D-P-G.

NAME: Uncharacterized protein family UPF0012 signature.
 CONSENSUS: [GTA]-x(2)-[IVT]-C-Y-D-[LIVM]-x-F-P-x(9)-G.

NAME: Uncharacterized protein family UPF0015 signature.
 CONSENSUS: [DE]-[LIVMF](3)-R-T-[SG]-G-x(2)-R-x-S-x-[FY]-[LIVM](2)-W-Q.

NAME: Uncharacterized protein family UPF0016 signature.
 CONSENSUS: E-[LIVM]-G-D-K-T-F-[LIVMF](2)-A.

NAME: Uncharacterized protein family UPF0017 signature.
 CONSENSUS: D-x(8)-[GN]-[LFY]-x(4)-[DET]-[LY]-Y-x(3)-[ST]-x(7)-[IV]-x(2)-[PS]-x-
 CONSENSUS: [LIVM]-x-[LIVM]-x(3)-[DN]-D.

NAME: Uncharacterized protein family UPF0019 signature.
 CONSENSUS: L-P-V-[VT]-[NQL]-F-[AT]-A-G-G-[LIV]-A-T-P-A-D-A-A-[LM].

NAME: Uncharacterized protein family UPF0020 signature.
 CONSENSUS: D-P-[LIVMF]-C-G-[ST]-G-x(3)-[LI]-E.

NAME: Uncharacterized protein family UPF0021 signature.
 CONSENSUS: C-K-x(2)-F-x(4)-E-x(22,23)-S-G-G-K-D.

NAME: Uncharacterized protein family UPF0023 signature.
 CONSENSUS: D-x-D-E-[LIV]-L-x(4)-V-F-x(3)-S-K-G.

NAME: Uncharacterized protein family UPF0024 signature.
 CONSENSUS: G-x-K-D-[KR]-x-A-[LV]-T-x-Q-x-[LIVF]-[SGC].

NAME: Uncharacterized protein family UPF0025 signature.
 CONSENSUS: D-V-[LIV]-x(2)-G-H-[ST]-H-x(12)-[LIVMF]-N-P-G.

NAME: Uncharacterized protein family UPF0027 signature.
 CONSENSUS: Q-[LIVM]-x-N-x-A-x-[LIVM]-P-x-I-x(6)-[LIVM]-P-D-x-H-x-G-x-G-x(2)-[IV]-G.

NAME: Uncharacterized protein family UPF0028 signature.

CONSENSUS: [GA]-[GS]-G-[GA]-A-R-G-x-[SA]-H-x-G-x(9)-[IV]-x-[IV]-D-x(2)-[GA]-G-x-S-
 CONSENSUS: x-G.

NAME: Uncharacterized protein family UPF0029 signature.

CONSENSUS: G-x(2)-[LIVM](2)-x(2)-[LIVM]-x(4)-[LIVM]-x(5)-[LIVM](2)-x-R-[FYW](2)-G-
 CONSENSUS: G-x(2)-[LIVM]-G.

NAME: Uncharacterized protein family UPF0030 signature.

CONSENSUS: [GA]-L-I-[LIV]-P-G-G-E-S-T-[STA].

NAME: Uncharacterized protein family UPF0031 signature 1.

CONSENSUS: [SAV]-[IVW]-[LVA]-[LIV]-G-[PNS]-G-L-[GP]-x-[DENQT].

NAME: Uncharacterized protein family UPF0031 signature 2.

CONSENSUS: [GA]-G-x-G-D-[TV]-[LT]-[STA]-G-x-[LIVM].

NAME: Uncharacterized protein family UPF0032 signature.

CONSENSUS: Y-x(2)-F-[LIVMA](2)-x-L-x(4)-G-x(2)-F-[EQ]-[LIVMF]-P-[LIVM].

NAME: Uncharacterized protein family UPF0033 signature.

CONSENSUS: L-[DN]-x(2)-[TAG]-x(2)-C-P-x-P-x-[LIVM].

NAME: Uncharacterized protein family UPF0034 signature.

CONSENSUS: [LIVM]-[DNG]-[LIVM]-N-x-G-C-P-x(3)-[LIVMASQ]-x(5)-G-[SAC].

NAME: Uncharacterized protein family UPF0035 signature.

CONSENSUS: L-L-T-x-R-[SA]-x(3)-R-x(3)-G-x(3)-F-P-G-G.

NAME: Uncharacterized protein family UPF0036 signature.

CONSENSUS: H-x-S-G-H-[GA]-x(3)-[DE]-x(3)-[LM]-x(5)-P-x(3)-[LIVM]-P-x-H-G-[DE].

NAME: Uncharacterized protein family UPF0038 signature.

CONSENSUS: G-x-[LI]-x-R-x(2)-L-x(4)-F-x(8)-[LIV]-x(5)-P-x-[LIV].

NAME: Uncharacterized protein family UPF0044 signature.

CONSENSUS: L-[ST]-x(3)-K-x(3)-[KR]-[SGA]-x-[GA]-H-x-L-x-P-[LIV]-x(2)-[LIV]-[GA]-
 CONSENSUS: x(2)-G.

NAME: Uncharacterized protein family UPF0047 signature.

CONSENSUS: S-X(2)-[LIV]-x-[LIV]-x(2)-G-x(4)-G-T-W-Q-x-[LIV].

NAME: Uncharacterized protein family UPF0054 signature.

CONSENSUS: H-[GS]-x-L-H-L-[LI]-G-[FYW]-D-H.

NAME: Uncharacterized protein family UPF0057 signature.

CONSENSUS: [LIV]-x-[STA]-[LIVF](3)-P-P-[LIVA]-[GA]-[IV]-x(4)-[GKN].

NAME: Hypothetical YER057c/yjjV family signature.

CONSENSUS: P-[AT]-R-[SA]-x-[LIVMY]-x(2)-[AK]-x-L-P-x(4)-[LIVM]-E.

NAME: Hypothetical hesB/yadR/yfhF family signature.

CONSENSUS: F-x-[LIVMFY]-x-N-[PG]-[NSK]-x(4)-C-x-C-[GS]-x-S-F.

NAME: Hypothetical yabO/yceC/sfhB family signature.

CONSENSUS: [NHY]-R-[LI]-D-x(2)-T-[ST]-G-[LIVMA]-[LIVMF](2)-[LIVMFG]-[SGAC].

We claim:

1. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_16c16; hfbr2_16f21; hfbr2_16g18; hfbr2_16i12; hfbr2_16k22; hfbr2_16l12; hfbr2_22f21; hfbr2_22h13; hfbr2_22h13; hfbr2_22i4; hfbr2_22k3; hfbr2_22k8; hfbr2_23b10; hfbr2_23b21; hfbr2_23f2; hfbr2_23l24; hfbr2_23n16; hfbr2_23o24; hfbr2_23o5; hfbr2_2a2; hfbr2_2b17; hfbr2_2b5; hfbr2_2c1; hfbr2_2c17; hfbr2_2c18; hfbr2_2d15; hfbr2_2d17; hfbr2_2d20; hfbr2_2g18; hfbr2_2h1; hfbr2_2h10; hfbr2_2i17; hfbr2_2k14; hfbr2_2k19; hfbr2_3b16; hfbr2_3c18; hfbr2_3f16; hfbr2_3g8; hfbr2_3l2; hfbr2_4l15; hfbr2_62b11; hfbr2_62f10; hfbr2_62l19; hfbr2_62n10; hfbr2_62o17; hfbr2_64a11; hfbr2_64a15; hfbr2_64c16; hfbr2_64c4; hfbr2_64h6; hfbr2_64i20; hfbr2_64j18; hfbr2_64k24; hfbr2_64o16; hfbr2_6a17; hfbr2_6b24; hfbr2_6i20; hfbr2_6o17; hfbr2_7l10; hfbr2_72b18; hfbr2_72d13; hfbr2_72l12; hfbr2_72m16; hfbr2_72n12; hfbr2_78c24; hfbr2_78d13; hfbr2_78k24; hfbr2_78n23; hfbr2_7a24; hfbr2_7e22; hfbr2_7j4; hfbr2_82c20; hfbr1_10c20; hfbr2_82e17; hfbr1_10e17; hfbr2_82e4;; hfbr1_10e4; hfbr2_82g14;; hfbr1_10g14; hfbr2_82i17;; hfbr1_10; hfbr2_82i24;; hfbr1_10; hfbr2_82m16;; hfbr1_10; hfbr2_82m6;; hfbr1_10; hfkd2_1j9; hfkd2_24a15; hfkd2_24b15; hfkd2_24e23; hfkd2_24n20; hfkd2_24p5; hfkd2_3i13; hfkd2_3o17; hfkd2_46a6; hfkd2_46b10; hfkd2_46d13; hfkd2_46j20; hfkd2_46k19; hfkd2_46m4; hfkd2_47a4; hfkd2_4b6; hfkd2_4c8; hfkd2_4k14; hfkd2_4m11; hmcfl_1a11; hmcfl_1c23; hmcfl_1e15; hmcfl_1g13; htes3_1n3; htes3_14g5; htes3_14h21; htes3_14p14; htes3_14p7; htes3_15a13; Htes3_15c24; htes3_15c6; htes3_15g14; htes3_15h1; htes3_15i5; htes3_15j18; Htes3_15j3; htes3_15k11; htes3_17f10; htes3_17l17; htes3_17n12; htes3_17n18; Htes3_18f3; htes3_18l7; htes3_19f19; htes3_19j17; htes3_1c1; htes3_1g13; htes3_1k11; htes3_20c21; htes3_20k2; htes3_20m18; htes3_21d4; htes3_21j15; htes3_21l16; htes3_21n23; htes3_22c23; htes3_22g2; htes3_22n13; htes3_23l11; htes3_23n19; Htes3_23n19; htes3_26g22; htes3_27d1; htes3_27k4; htes3_27o14; htes3_28d14; htes3_2a11; htes3_2a17; htes3_2d15; htes3_2e12; htes3_2f14; htes3_2g7; htes3_2h1; htes3_2h15; htes3_2l19; htes3_2m18; htes3_2m20; htes3_2n9; htes3_2ol3; htes3_30f4; Htes3_35b4; htes3_35b5; htes3_35e21; htes3_35g6; htes3_35k16; htes3_35k24; htes3_35n12; htes3_35n24; htes3_35n9; htes3_35p17; htes3_35p22; htes3_4b4; htes3_4f17; htes3_4f5; htes3_4h6; htes3_4o19; htes3_50j4; htes3_50n06;

htes3_50n23; htes3_6b21; htes3_6c11; htes3_6d16; htes3_72k11; Htes3_72k15;
 htes3_72p16; htes3_7b22; htes3_7d17; htes3_7j3; htes3_7j8; htes3_7p10; htes3_7p9;
 htes3_8e24; Htes3_8g11; Htes3_8g5; htes3_8m10; Htes3_8p7; Htes3_9e22; Htes3_9i20;
 Htes3_9k22; hute1_17k7; hute1_18c12; hute1_18i19; hute1_18i4; hute1_18l1;
 hute1_19f19; hute1_19g19; hute1_19g22; hute1_19h17; hute1_19j11; hute1_1i2;
 hute1_20b19; hute1_20g21; hute1_20h13; hute1_20m11; hute1_20m24; hute1_21d15;
 hute1_22d2; hute1_22e12; hute1_22n2; hute1_22o2; hute1_23e13; hute1_23g11;
 hute1_24c19; hute1_24e11; hute1_24j6; hute1_2h3; their complements; and variants thereof.

2. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_16c16; hfbr2_16f21; hfbr2_16g18; hfbr2_16i12; hfbr2_16k22; hfbr2_16l12; hfbr2_22f21; hfbr2_22h13; hfbr2_22h13; hfbr2_22i4; hfbr2_22k3; hfbr2_22k8; hfbr2_23b10; hfbr2_23b21; hfbr2_23f2; hfbr2_23l24; ; hfbr2_23n16; hfbr2_23o24; hfbr2_23o5; hfbr2_2a2; hfbr2_2b17; hfbr2_2b5; hfbr2_2c1; hfbr2_2c17; hfbr2_2c18; hfbr2_2d15; hfbr2_2d17; hfbr2_2d20; hfbr2_2g18; hfbr2_2h1; hfbr2_2h10; hfbr2_2i17; hfbr2_2k14; hfbr2_2k19; hfbr2_3c18; hfbr2_3f16; hfbr2_3g8; hfbr2_3l2; hfbr2_4l15; hfbr2_62b11; hfbr2_62f10; hfbr2_62l19; hfbr2_62n10; hfbr2_62o17; hfbr2_64a11; hfbr2_64a15; hfbr2_64c16; hfbr2_64c4; hfbr2_64h6; hfbr2_64i20; hfbr2_64j18; hfbr2_64k24; hfbr2_64o16; hfbr2_6a17; hfbr2_6b24; hfbr2_6i20; hfbr2_6o17; hfbr2_7l10; hfbr2_72b18; hfbr2_72d13; hfbr2_72l12; hfbr2_72m16; hfbr2_72n12; hfbr2_78c24; hfbr2_78d13; hfbr2_78k24; hfbr2_78n23; hfbr2_7a24; hfbr2_7e22; hfbr2_7j4; hfbr2_82c20; hfbr1_10c20; hfbr2_82e17; hfbr1_10e17; hfbr2_82e4; hfbr1_10e4; hfbr2_82g14; hfbr1_10g14; hfbr2_82i17; hfbr1_10; hfbr2_82i24; hfbr1_10; hfbr2_82m16; hfbr1_10; hfbr2_82m6; hfbr1_10; their complements; and variants thereof.

3. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_16f21; hfbr2_16k22; hfbr2_22f21; hfbr2_22h13; hfbr2_22i4; hfbr2_22k3; hfbr2_22k8; hfbr2_23f2; ; hfbr2_23o24; hfbr2_23o5; hfbr2_2a2; hfbr2_2c1; hfbr2_2c18; hfbr2_2d20; hfbr2_2g18; hfbr2_2h1; hfbr2_2h10; hfbr2_2k19; hfbr2_3f16; hfbr2_3l2; hfbr2_62n10; hfbr2_64a11; hfbr2_64c16; hfbr2_64c4; hfbr2_64h6; hfbr2_64i20; hfbr2_64o16; hfbr2_6a17; hfbr2_6i20; hfbr2_7l10;

hfbr2_72d13; hfbr2_72m16; hfbr2_72n12; hfbr2_78d13; hfbr2_78n23; hfbr2_7a24;
hfbr2_7e22; hfbr2_7j4; hfbr2_82m16; and hfbr1_10.

4. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfkd2_1j9; hfkd2_24a15; hfkd2_24b15; hfkd2_24e23; hfkd2_24n20; hfkd2_24p5; hfkd2_3i13; hfkd2_3o17; hfkd2_46a6; hfkd2_46b10; hfkd2_46d13; hfkd2_46j20; hfkd2_46k19; hfkd2_46m4; hfkd2_47a4; hfkd2_4b6; hfkd2_4c8; hfkd2_4k14; hfkd2_4m11; their complements; and variants thereof.

5. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfkd2_1j9; hfkd2_24e23; hfkd2_46a6; hfkd2_46b10; hfkd2_46d13; hfkd2_4b6; hfkd2_4c8; their complements; and variants thereof.

6. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hmcfl_1a11; hmcfl_1c23; hmcfl_1e15; hmcfl_1g13; their complements; and variants thereof.

7. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hmcfl_1c23 hmcfl_1g13; their complements; and variants thereof.

8. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: htes3_1n3; htes3_14g5; htes3_14h21; htes3_14p14; htes3_14p7; htes3_15a13; Htes3_15c24; htes3_15c6; htes3_15g14; htes3_15h1; htes3_15i5; htes3_15j18; Htes3_15j3; htes3_15k11; htes3_17f10; htes3_17i17; htes3_17n12; htes3_17n18; Htes3_18f3; htes3_18i7; htes3_19f19; htes3_19j17; htes3_1c1; htes3_1g13; htes3_1k11; htes3_20c21; htes3_20k2; htes3_20m18; htes3_21d4; htes3_21j15; htes3_21i16; htes3_21n23; htes3_22c23; htes3_22g2; htes3_22n13; htes3_23i11; htes3_23n19; Htes3_23n19; htes3_26g22; htes3_27d1; htes3_27k4; htes3_27o14; htes3_28d14; htes3_2a11; htes3_2a17; htes3_2d15; htes3_2e12; htes3_2f14; htes3_2g7; htes3_2h1; htes3_2h15; htes3_2i19; htes3_2m18; htes3_2m20; htes3_2n9; htes3_2o13; htes3_30f4; Htes3_35b4; htes3_35b5; htes3_35e21;

htes3_35g6; htes3_35k16; htes3_35k24; htes3_35n12; htes3_35n24; htes3_35n9;
htes3_35p17; htes3_35p22; htes3_4b4; htes3_4f17; htes3_4f5; htes3_4h6; htes3_4o19;
htes3_50j4; htes3_50n06; htes3_50n23; htes3_6b21; htes3_6c11; htes3_6d16; htes3_72k11;
Htes3_72k15; htes3_72p16; htes3_7b22; htes3_7d17; htes3_7j3; htes3_7j8; htes3_7p10;
htes3_7p9; htes3_8e24; Htes3_8g11; Htes3_8g5; htes3_8m10; Htes3_8p7; Htes3_9e22;
Htes3_9i20; Htes3_9k22; their complements; and variants thereof.

9. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: htes3_14g5; htes3_14p14; htes3_14p7; htes3_15a13; htes3_15g14; htes3_15h1; htes3_15j18; htes3_17f10; Htes3_18f3; htes3_19f19; htes3_19j17; htes3_20c21; htes3_21n23; htes3_22c23; htes3_22n13; Htes3_23n19; htes3_27o14; htes3_28d14; htes3_2a11; htes3_2d15; htes3_2f14; htes3_2g7; htes3_2h15; htes3_2i19; htes3_2m20; htes3_2n9; htes3_30f4; htes3_35g6; htes3_35n24; htes3_35p17; htes3_4b4; htes3_4f17; htes3_4o19; htes3_50j4; htes3_50n23; htes3_50n06; htes3_6b21; htes3_6d16; htes3_72k11; htes3_7d17; htes3_7j8; Htes3_8g11; Htes3_8g5; Htes3_8p7; Htes3_9e22; Htes3_9i20; Htes3_9k22; their complements; and variants thereof.

10. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_16g18; hfbr2_2k14; Htes3_35b4; htes3_35p22; htes3_7j3; htes3_7p10; hute1_20m11; their complements; and variants thereof.

11. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_16c16; hfbr2_2b5; htes3_15i5; htes3_18i7; htes3_1k11; Htes3_72k15; htes3_7b22; hute1_19g22; hute1_24j6; their complements; and variants thereof.

12. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_2d15; htes3_35e21; hute1_2h3; their complements; and variants thereof.

13. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_23l24; hfbr2_2i17; hfbr2_41m15; hfbr2_62f10; hfbr2_62l19; hfbr2_64j18;

hfkd2_24n20; hfkd2_24p5; hfkd2_4k14; htes3_1g13; htes3_21l16; htes3_23l11; htes3_26g22; htes3_4h6; htes3_72p16; hute1_19h17; hute1_20h13; hute1_24e11; their complements; and variants thereof.

14. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_3g8; hfbr2_62o17; hfbr2_6b24; hfbr2_78k24; hfkd2_24b15; hfkd2_3o17; hfkd2_46j20; htes3_17l17; htes3_17n18; htes3_27d1; htes3_2a17; htes3_35b5; htes3_35k16; htes3_35n12; htes3_35n9; hute1_20b19; hute1_20m24; hute1_23e13; their complements; and variants thereof.

15. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_23b10; hfbr2_3c18; hfbr2_64a15; hfbr2_6o17; hfbr2_72b18; hfbr2_72l12; hfbr2_82i24(hfbr1_10); htes3_14h21; Htes3_15j3; htes3_20m18; htes3_22g2; htes3_2m18; htes3_7p9; htes3_8m10; hute1_18l1; their complements; and variants thereof.

16. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_23b21; hfbr2_23n16; hfbr2_2c17; hfbr2_62b11; hfbr2_78c24; hfbr2_82e4 (hfbr1_10e4); hfbr2_82i17 (hfbr1_10); hfbr2_82m6 (hfbr1_10); hfkd2_46m4; htes3_15k11; htes3_1c1; hhtes3_1n3; htes3_20k2; htes3_21d4; htes3_23n19; htes3_4f5; htes3_6c11; htes3_8e24; hute1_20g21; hute1_22d2; hute1_22e12; their complements; and variants thereof.

17. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_16i12; hfbr2_16l12; hfbr2_22h13; hfbr2_2b17; hfbr2_2d17; hfbr2_64k24; hfbr2_82c20 (hfbr1_10c20); hfbr2_82e17 (hfbr1_10e17); hfbr2_82g14 (hfbr1_10g14); hfkd2_24a15; hfkd2_3i13; hfkd2_4m11; hmcfl_1a11; hmcfl_1e15; htes3_15c6; htes3_2ol3; htes3_27k4; htes3_2h1; htes3_35k24; hute1_19f19; and hute1_24c19; their complements; and variants thereof.

18. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfkd2_46k19; hfkd2_47a4;

htes3_2e12; htes3_21j15; htes3_17n12; hute1_18i19; hute1_1i2; their complements; and variants thereof.

19. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hute1_17k7; hute1_18c12; hute1_18i19; hute1_18i4; hute1_18l1; hute1_19f19; hute1_19g19; hute1_19g22; hute1_19h17; hute1_19j11; hute1_1i2; hute1_20b19; hute1_20g21; hute1_20h13; hute1_20m11; hute1_20m24; hute1_21d15; hute1_22d2; hute1_22e12; hute1_22n2; hute1_22o2; hute1_23e13; hute1_23g11; hute1_24c19; hute1_24e11; hute1_24j6; hute1_2h3; their complements; and variants thereof.

20. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hute1_17k7; hute1_18c12; hute1_18i4; hute1_19g19; hute1_19j11; hute1_22n2; hute1_21d15; hute1_22o2; hute1_23g11; their complements; and variants thereof.

21. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfbr2_16c16; hfbr2_16f21; hfbr2_16g18; hfbr2_16i12; hfbr2_16k22; hfbr2_16l12; hfbr2_22f21; hfbr2_22h13; hfbr2_22h13; hfbr2_22i4; hfbr2_22k3; hfbr2_22k8; hfbr2_23b10; hfbr2_23b21; hfbr2_23f2; hfbr2_23l24; ; hfbr2_23n16; hfbr2_23o24; hfbr2_23o5; hfbr2_2a2; hfbr2_2b17; hfbr2_2b5; hfbr2_2c1; hfbr2_2c17; hfbr2_2c18; hfbr2_2d15; hfbr2_2d17; hfbr2_2d20; hfbr2_2g18; hfbr2_2h1; hfbr2_2h10; hfbr2_2i17; hfbr2_2k14; hfbr2_2k19; hfbr2_3c18; hfbr2_3f16; hfbr2_3g8; hfbr2_3l2; hfbr2_4l15; hfbr2_62b11; hfbr2_62f10; hfbr2_62l19; hfbr2_62n10; hfbr2_62o17; hfbr2_64a11; hfbr2_64a15; hfbr2_64c16; hfbr2_64c4; hfbr2_64h6; hfbr2_64i20; hfbr2_64j18; hfbr2_64k24; hfbr2_64o16; hfbr2_6a17; hfbr2_6b24; hfbr2_6i20; hfbr2_6o17; hfbr2_71o20; hfbr2_72b18; hfbr2_72d13; hfbr2_72l12; hfbr2_72m16; hfbr2_72n12; hfbr2_78c24; hfbr2_78d13; hfbr2_78k24; hfbr2_78n23; hfbr2_7a24; hfbr2_7e22; hfbr2_7j4; hfbr2_82c20; hfbr1_10c20; hfbr2_82e17; hfbr1_10e17; hfbr2_82e4;; hfbr1_10e4; hfbr2_82g14;; hfbr1_10g14; hfbr2_82i17;; hfbr1_10; hfbr2_82i24;; hfbr1_10; hfbr2_82m16;; hfbr1_10; hfbr2_82m6;; hfbr1_10; hfkd2_1j9; hfkd2_24a15; hfkd2_24b15; hfkd2_24e23; hfkd2_24n20; hfkd2_24p5; hfkd2_3i13; hfkd2_3o17; hfkd2_46a6;

hfk2_46b10; hfk2_46d13; hfk2_46j20; hfk2_46k19; hfk2_46m4; hfk2_47a4;
hfk2_4b6; hfk2_4c8; hfk2_4k14; hfk2_4m11; hmc1_1a11; hmc1_1c23; hmc1_1e15;
hmc1_1g13; hhtes3_1n3; htes3_14g5; htes3_14h21; htes3_14p14; htes3_14p7;
htes3_15a13; Htes3_15c24; htes3_15c6; htes3_15g14; htes3_15h1; htes3_15i5;
htes3_15j18; Htes3_15j3; htes3_15k11; htes3_17f10; htes3_17i17; htes3_17n12;
htes3_17n18; Htes3_18f3; htes3_18i7; htes3_19f19; htes3_19j17; htes3_1c1; htes3_1g13;
htes3_1k11; htes3_20c21; htes3_20k2; htes3_20m18; htes3_21d4; htes3_21j15;
htes3_21i16; htes3_21n23; htes3_22c23; htes3_22g2; htes3_22n13; htes3_23i11;
htes3_23n19; Htes3_23n19; htes3_26g22; htes3_27d1; htes3_27k4; htes3_27o14;
htes3_28d14; htes3_2a11; htes3_2a17; htes3_2d15; htes3_2e12; htes3_2f14; htes3_2g7;
htes3_2h1; htes3_2h15; htes3_2i19; htes3_2m18; htes3_2m20; htes3_2n9; htes3_2o13;
htes3_30f4; Htes3_35b4; htes3_35b5; htes3_35e21; htes3_35g6; htes3_35k16;
htes3_35k24; htes3_35n12; htes3_35n24; htes3_35n9; htes3_35p17; htes3_35p22;
htes3_4b4; htes3_4f17; htes3_4f5; htes3_4h6; htes3_4o19; htes3_50j4; htes3_50n06;
htes3_50n23; htes3_6b21; htes3_6c11; htes3_6d16; htes3_72k11; Htes3_72k15;
htes3_72p16; htes3_7b22; htes3_7d17; htes3_7j3; htes3_7j8; htes3_7p10; htes3_7p9;
htes3_8e24; Htes3_8g11; Htes3_8g5; htes3_8m10; Htes3_8p7; Htes3_9e22; Htes3_9i20;
Htes3_9k22; hute1_17k7; hute1_18c12; hute1_18i19; hute1_18i4; hute1_18i1;
hute1_19f19; hute1_19g19; hute1_19g22; hute1_19h17; hute1_19j11; hute1_1i2;
hute1_20b19; hute1_20g21; hute1_20h13; hute1_20m11; hute1_20m24; hute1_21d15;
hute1_22d2; hute1_22e12; hute1_22n2; hute1_22o2; hute1_23e13; hute1_23g11;
hute1_24c19; hute1_24e11; hute1_24j6; hute1_2h3; their complements; and variants thereof.

22. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of:

hfbr2_16c16; hfbr2_16f21; hfbr2_16g18; hfbr2_16i12; hfbr2_16k22; hfbr2_16l12;
hfbr2_22f21; hfbr2_22h13; hfbr2_22h13; hfbr2_22i4; hfbr2_22k3; hfbr2_22k8;
hfbr2_23b10; hfbr2_23b21; hfbr2_23f2; hfbr2_23i24; ; hfbr2_23n16; hfbr2_23o24;
hfbr2_23o5; hfbr2_2a2; hfbr2_2b17; hfbr2_2b5; hfbr2_2c1; hfbr2_2c17; hfbr2_2c18;
hfbr2_2d15; hfbr2_2d17; hfbr2_2d20; hfbr2_2g18; hfbr2_2h1; hfbr2_2h10; hfbr2_2i17;
hfbr2_2k14; hfbr2_2k19; hfbr2_3c18; hfbr2_3f16; hfbr2_3g8; hfbr2_3i2; hfbr2_41m15;
hfbr2_62b11; hfbr2_62f10; hfbr2_62i19; hfbr2_62n10; hfbr2_62o17; hfbr2_64a11;

hfbr2_64a15; hfbr2_64c16; hfbr2_64c4; hfbr2_64h6; hfbr2_64i20; hfbr2_64j18;
 hfbr2_64k24; hfbr2_64o16; hfbr2_6a17; hfbr2_6b24; hfbr2_6i20; hfbr2_6o17;
 hfbr2_71o20; hfbr2_72b18; hfbr2_72d13; hfbr2_72l12; hfbr2_72m16; hfbr2_72n12;
 hfbr2_78c24; hfbr2_78d13; hfbr2_78k24; hfbr2_78n23; hfbr2_7a24; hfbr2_7e22;
 hfbr2_7j4; hfbr2_82c20; hfbr1_10c20; hfbr2_82e17; hfbr1_10e17; hfbr2_82e4;
 hfbr1_10e4; hfbr2_82g14; hfbr1_10g14; hfbr2_82i17; hfbr1_10; hfbr2_82i24; hfbr1_10;
 hfbr2_82m16; hfbr1_10; hfbr2_82m6; hfbr1_10; complements of the nucleic acid
 sequences; and variants thereof.

23. A computer readable medium, comprising in electronic form at least one
 nucleic acid or protein sequence of a clone selected from the group consisting of:
 hfbr2_16f21; hfbr2_16k22; hfbr2_22f21; hfbr2_22h13; hfbr2_22i4; hfbr2_22k3; hfbr2_22k8;
 hfbr2_23f2; ; hfbr2_23o24; hfbr2_23o5; hfbr2_2a2; hfbr2_2c1; hfbr2_2c18; hfbr2_2d20;
 hfbr2_2g18; hfbr2_2h1; hfbr2_2h10; hfbr2_2k19; hfbr2_3f16; hfbr2_3l2; hfbr2_62n10;
 hfbr2_64a11; hfbr2_64c16; hfbr2_64c4; hfbr2_64h6; hfbr2_64i20; hfbr2_64k24;
 hfbr2_64o16; hfbr2_6a17; hfbr2_6i20; hfbr2_71o20; hfbr2_72d13; hfbr2_72m16;
 hfbr2_72n12; hfbr2_78d13; hfbr2_78n23; hfbr2_7a24; hfbr2_7e22; hfbr2_7j4; hfbr2_82m16;
 hfbr1_10; complements of the nucleic acid sequences; and variants thereof.

24. A computer readable medium, comprising in electronic form at least one
 nucleic acid or protein sequence of a clone selected from the group consisting of:
 hfkd2_1j9; hfkd2_24a15; hfkd2_24b15; hfkd2_24e23; hfkd2_24n20; hfkd2_24p5;
 hfkd2_3i13; hfkd2_3o17; hfkd2_46a6; hfkd2_46b10; hfkd2_46d13; hfkd2_46j20;
 hfkd2_46k19; hfkd2_46m4; hfkd2_47a4; hfkd2_4b6; hfkd2_4c8; hfkd2_4k14;
 hfkd2_4m11; complements of the nucleic acid sequences; and variants thereof.

25. A computer readable medium, comprising in electronic form at least one
 nucleic acid or protein sequence of a clone selected from the group consisting of: hfkd2_1j9;
 hfkd2_24e23; hfkd2_46a6; hfkd2_46b10; hfkd2_46d13; hfkd2_4b6; hfkd2_4c8;
 complements of the nucleic acid sequences; and variants thereof.

26. A computer readable medium, comprising in electronic form at least one
 nucleic acid or protein sequence of a clone selected from the group consisting of:

hmcfl_1a11; hmcfl_1c23; hmcfl_1e15; hmcfl_1g13; complements of the nucleic acid sequences; and variants thereof.

27. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hmcfl_1c23; hmcfl_1g13; complements of the nucleic acid sequences; and variants thereof.

28. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hhtes3_1n3; htes3_14g5; htes3_14h21; htes3_14p14; htes3_14p7; htes3_15a13; Htes3_15c24; htes3_15c6; htes3_15g14; htes3_15h1; htes3_15i5; htes3_15j18; Htes3_15j3; htes3_15k11; htes3_17f10; htes3_17l17; htes3_17n12; htes3_17n18; Htes3_18f3; htes3_18l7; htes3_19f19; htes3_19j17; htes3_1c1; htes3_1g13; htes3_1k11; htes3_20c21; htes3_20k2; htes3_20m18; htes3_21d4; htes3_21j15; htes3_21l16; htes3_21n23; htes3_22c23; htes3_22g2; htes3_22n13; htes3_23l11; htes3_23n19; Htes3_23n19; htes3_26g22; htes3_27d1; htes3_27k4; htes3_27o14; htes3_28d14; htes3_2a11; htes3_2a17; htes3_2d15; htes3_2e12; htes3_2f14; htes3_2g7; htes3_2h1; htes3_2h15; htes3_2l19; htes3_2m18; htes3_2m20; htes3_2n9; htes3_2ol3; htes3_30f4; Htes3_35b4; htes3_35b5; htes3_35e21; htes3_35g6; htes3_35k16; htes3_35k24; htes3_35n12; htes3_35n24; htes3_35n9; htes3_35p17; htes3_35p22; htes3_4b4; htes3_4f17; htes3_4f5; htes3_4h6; htes3_4o19; htes3_50j4; htes3_50n06; htes3_50n23; htes3_6b21; htes3_6c11; htes3_6d16; htes3_72k11; Htes3_72k15; htes3_72p16; htes3_7b22; htes3_7d17; htes3_7j3; htes3_7j8; htes3_7p10; htes3_7p9; htes3_8e24; Htes3_8g11; Htes3_8g5; htes3_8m10; Htes3_8p7; Htes3_9e22; Htes3_9i20; Htes3_9k22; complements of the nucleic acid sequences; and variants thereof.

29. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: htes3_14g5; htes3_14p14; htes3_14p7; htes3_15a13; htes3_15g14; htes3_15h1; htes3_15j18; htes3_17f10; htes3_17n18; Htes3_18f3; htes3_19f19; htes3_19j17; htes3_20c21; htes3_21n23; htes3_22c23; htes3_22n13; Htes3_23n19; htes3_27o14; htes3_28d14; htes3_2a11; htes3_2d15; htes3_2f14; htes3_2g7; htes3_2h15; htes3_2l19; htes3_2m20; htes3_2n9; htes3_30f4; htes3_35g6; htes3_35n24; htes3_35p17; htes3_4b4; htes3_4f17;

htes3_4o19; htes3_50j4; htes3_50n23; htes3_50n06; htes3_6b21; htes3_6d16; htes3_72k11; htes3_7d17; htes3_7j8; Htes3_8g11; Htes3_8g5; Htes3_8p7; Htes3_9e22; Htes3_9i20; Htes3_9k22; complements of the nucleic acid sequences; and variants thereof.

30. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfbr2_16g18; hfbr2_2k14; Htes3_35b4; htes3_35p22; htes3_7j3; htes3_7p10; hute1_20m11; complements of the nucleic acid sequences; and variants thereof.

31. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfbr2_16c16; hfbr2_2b5; htes3_15i5; htes3_18l7; htes3_1k11; Htes3_72k15; htes3_7b22; hute1_19g22; hute1_24j6; complements of the nucleic acid sequences; and variants thereof.

32. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfbr2_2d15; htes3_35e21; hute1_2h3; complements of the nucleic acid sequences; and variants thereof.

33. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfbr2_23l24; hfbr2_2i17; hfbr2_41m15; hfbr2_62f10; hfbr2_62l19; hfbr2_64j18; hfkd2_24n20; hfkd2_24p5; hfkd2_4k14; htes3_1g13; htes3_21l16; htes3_23l11; htes3_26g22; htes3_4h6; htes3_72p16; hute1_19h17; hute1_20h13; hute1_24e11; complements of the nucleic acid sequences; and variants thereof.

34. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfbr2_3g8; hfbr2_62o17; hfbr2_6b24; hfbr2_78k24; hfkd2_24b15; hfkd2_3o17; hfkd2_46j20; htes3_17l17; Htes3_17n18; htes3_27d1; htes3_2a17; htes3_35b5; htes3_35k16; htes3_35n12; htes3_35n9; hute1_20b19; hute1_20m24; hute1_23e13; complements of the nucleic acid sequences; and variants thereof.

35. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of:

hfbr2_23b10; hfbr2_3c18; hfbr2_64a15; hfbr2_6o17; hfbr2_72b18; hfbr2_72i12;
hfbr2_82i24(hfbr1_10); htes3_14h21; Htes3_15j3; htes3_20m18; htes3_22g2; htes3_2m18;
htes3_7p9; htes3_8m10; hute1_18i1; complements of the nucleic acid sequences; and
variants thereof.

36. A computer readable medium, comprising in electronic form at least one
nucleic acid or protein sequence of a clone selected from the group consisting of:
hfbr2_23b21; hfbr2_23n16; hfbr2_2c17; hfbr2_62b11; hfbr2_78c24; hfbr2_82e4
(hfbr1_10e4); hfbr2_82i17 (hfbr1_10); hfbr2_82m6 (hfbr1_10); hfkd2_46m4; htes3_15k11;
htes3_1c1; hhtes3_1n3; htes3_20k2; htes3_21d4; htes3_23n19; htes3_4f5; htes3_6c11;
htes3_8e24; hute1_20g21; hute1_22d2; hute1_22e12; complements of the nucleic acid
sequences; and variants thereof.

37. A computer readable medium, comprising in electronic form at least one
nucleic acid or protein sequence of a clone selected from the group consisting of:
hfbr2_16i12; hfbr2_16l12; hfbr2_22h13; hfbr2_2b17; hfbr2_2d17; hfbr2_64k24;
hfbr2_82c20 (hfbr1_10c20); hfbr2_82e17 (hfbr1_10e17); hfbr2_82g14 (hfbr1_10g14);
hfkd2_24a15; hfkd2_3i13; hfkd2_4m11; hmcfl_1a11; hmcfl_1e15; htes3_15c6;
htes3_2ol3; htes3_27k4; htes3_2h1; htes3_35k24; hute1_19f19; and hute1_24c19;
complements of the nucleic acid sequences; and variants thereof.

38. A computer readable medium, comprising in electronic form at least one
nucleic acid or protein sequence of a clone selected from the group consisting of:
hfkd2_46k19; hfkd2_47a4; htes3_2e12; htes3_21j15; htes3_17n12; hute1_18i19;
hute1_1i2; complements of the nucleic acid sequences; and variants thereof.

39. A computer readable medium, comprising in electronic form at least one
nucleic acid or protein sequence of a clone selected from the group consisting of:
hute1_17k7; hute1_18c12; hute1_18i19; hute1_18i4; hute1_18i1; hute1_19f19;
hute1_19g19; hute1_19g22; hute1_19h17; hute1_19j11; hute1_1i2; hute1_20b19;
hute1_20g21; hute1_20h13; hute1_20m11; hute1_20m24; hute1_21d15; hute1_22d2;
hute1_22e12; hute1_22n2; hute1_22o2; hute1_23e13; hute1_23g11; hute1_24c19;
hute1_24e11; hute1_24j6; hute1_2h3; complements of the nucleic acid sequences; and
variants thereof.

40. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of:
 hute1_17k7; hute1_18c12; hute1_18i4; hute1_19g19; hute1_19j11; hute1_22n2;
 hute1_21d15; hute1_22o2; hute1_23g11; complements of the nucleic acid sequences; and
 variants thereof.

41. A nucleic acid molecule having the sequence of a clone selected from the group consisting of hfbr2_16c16; hfbr2_16f21; hfbr2_16g18; hfbr2_16i12; hfbr2_16k22;
 hfbr2_16l12; hfbr2_22f21; hfbr2_22h13; hfbr2_22h13; hfbr2_22i4; hfbr2_22k3;
 hfbr2_22k8; hfbr2_23b10; hfbr2_23b21; hfbr2_23f2; hfbr2_23l24; hfbr2_23n16;
 hfbr2_23o24; hfbr2_23o5; hfbr2_2a2; hfbr2_2b17; hfbr2_2b5; hfbr2_2c1; hfbr2_2c17;
 hfbr2_2c18; hfbr2_2d15; hfbr2_2d17; hfbr2_2d20; hfbr2_2g18; hfbr2_2h1; hfbr2_2h10;
 hfbr2_2i17; hfbr2_2k14; hfbr2_2k19; hfbr2_3b16; hfbr2_3c18; hfbr2_3f16; hfbr2_3g8;
 hfbr2_3l2; hfbr2_4l1m15; hfbr2_62b11; hfbr2_62f10; hfbr2_62l19; hfbr2_62n10;
 hfbr2_62o17; hfbr2_64a11; hfbr2_64a15; hfbr2_64c16; hfbr2_64c4; hfbr2_64h6;
 hfbr2_64i20; hfbr2_64j18; hfbr2_64k24; hfbr2_64o16; hfbr2_6a17; hfbr2_6b24;
 hfbr2_6i20; hfbr2_6o17; hfbr2_71o20; hfbr2_72b18; hfbr2_72d13; hfbr2_72l12;
 hfbr2_72m16; hfbr2_72n12; hfbr2_78c24; hfbr2_78d13; hfbr2_78k24; hfbr2_78n23;
 hfbr2_7a24; hfbr2_7e22; hfbr2_7j4; hfbr2_82c20; hfbr1_10c20; hfbr2_82e17;
 hfbr1_10e17; hfbr2_82e4;; hfbr1_10e4; hfbr2_82g14;; hfbr1_10g14; hfbr2_82i17;;
 hfbr1_10; hfbr2_82i24;; hfbr1_10; hfbr2_82m16;; hfbr1_10; hfbr2_82m6;; hfbr1_10;
 hfkd2_1j9; hfkd2_24a15; hfkd2_24b15; hfkd2_24e23; hfkd2_24n20; hfkd2_24p5;
 hfkd2_3i13; hfkd2_3o17; hfkd2_46a6; hfkd2_46b10; hfkd2_46d13; hfkd2_46j20;
 hfkd2_46k19; hfkd2_46m4; hfkd2_47a4; hfkd2_4b6; hfkd2_4c8; hfkd2_4k14;
 hfkd2_4m11; hmcfl_1a11; hmcfl_1c23; hmcfl_1e15; hmcfl_1g13; htes3_1n3;
 htes3_14g5; htes3_14h21; htes3_14p14; htes3_14p7; htes3_15a13; Htes3_15c24;
 htes3_15c6; htes3_15g14; htes3_15h1; htes3_15i5; htes3_15j18; Htes3_15j3; htes3_15k11;
 htes3_17f10; htes3_17l17; htes3_17n12; htes3_17n18; Htes3_18f3; htes3_18l7;
 htes3_19f19; htes3_19j17; htes3_1c1; htes3_1g13; htes3_1k11; htes3_20c21; htes3_20k2;
 htes3_20m18; htes3_21d4; htes3_21j15; htes3_21l16; htes3_21n23; htes3_22c23;
 htes3_22g2; htes3_22n13; htes3_23l11; htes3_23n19; Htes3_23n19; htes3_26g22;
 htes3_27d1; htes3_27k4; htes3_27o14; htes3_28d14; htes3_2a11; htes3_2a17; htes3_2d15;
 htes3_2e12; htes3_2f14; htes3_2g7; htes3_2h1; htes3_2h15; htes3_2l19; htes3_2m18;

htes3_2m20; htes3_2n9; htes3_2ol3; htes3_30f4; Htes3_35b4; htes3_35b5; htes3_35e21; htes3_35g6; htes3_35k16; htes3_35k24; htes3_35n12; htes3_35n24; htes3_35n9; htes3_35p17; htes3_35p22; htes3_4b4; htes3_4f17; htes3_4f5; htes3_4h6; htes3_4o19; htes3_50j4; htes3_50n06; htes3_50n23; htes3_6b21; htes3_6c11; htes3_6d16; htes3_72k11; Htes3_72k15; htes3_72p16; htes3_7b22; htes3_7d17; htes3_7j3; htes3_7j8; htes3_7p10; htes3_7p9; htes3_8e24; Htes3_8g11; Htes3_8g5; htes3_8m10; Htes3_8p7; Htes3_9e22; Htes3_9i20; Htes3_9k22; hute1_17k7; hute1_18c12; hute1_18i19; hute1_18i4; hute1_18l1; hute1_19f19; hute1_19g19; hute1_19g22; hute1_19h17; hute1_19j11; hute1_1i2; hute1_20b19; hute1_20g21; hute1_20h13; hute1_20m11; hute1_20m24; hute1_21d15; hute1_22d2; hute1_22e12; hute1_22n2; hute1_22o2; hute1_23e13; hute1_23g11; hute1_24c19; hute1_24e11; hute1_24j6; hute1_2h3; their complements; and variants thereof.

42. A polypeptide encoded by the nucleic acid molecule according to claim 41.

43. An antibody or fragment thereof that is capable of binding to a specific portion of the peptide according to claim 42.

44. A pharmaceutical composition, comprising (a) an effective amount of a pharmaceutical agent, wherein said pharmaceutical agent is selected from the group consisting of the polypeptide according to claim 42, variants or functional derivatives thereof, and antibodies thereto; and (2) a physiologically acceptable carrier or excipient.

45. An expression vector comprising the nucleic acid molecule of claim 41 or a fragment thereof, and optionally a promoter operably linked to said nucleic acid molecule or said fragment.

46. A method for recombinantly producing a desired peptide, comprising expressing in a host cell a peptide encoded by the nucleic acid molecule according to claim 41.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
22 February 2001 (22.02.2001)

PCT

(10) International Publication Number
WO 01/12659 A3

(51) International Patent Classification⁷: **C12N 15/12**,
C07K 14/47, C12Q 1/68, C07K 16/18, A61K 38/17, C12P
21/00

(21) International Application Number: PCT/IB00/01496

(22) International Filing Date: 18 August 2000 (18.08.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/149,499 18 August 1999 (18.08.1999) US
60/156,503 28 September 1999 (28.09.1999) US

(63) Related by continuation (CON) or continuation-in-part
(CIP) to earlier application:
US 60/156,503 (CIP)
Filed on 18 August 1999 (18.08.1999)

(71) Applicant (for all designated States except US): **FRAUN-
HOFER-GESELLSCHAFT ZUR FOERDERUNG
DER ANGEWANDTEN FORSCHUNG E.V.** [DE/DE];
Leonrodstrasse 54, D-80636 München (DE).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **WIEMANN**,
Stefan [DE/DE]; Grosse Lachstrasse 30a, 69207 Sand-
hausen (DE). **POUSTKA**, Annemarie [DE/DE]; Laden-
burgstrasse 41, 69120 Heidelberg (DE). **WELLEN-
REUTHER**, Ruth [DE/DE]; An der Markscheide 5,
69126 Heidelberg (DE). **BLUM**, Helmut [DE/DE];
Koenigswieser Strasse 94, 81475 Muenchen (DE). **OBER-
MAIER**, Brigitte [DE/DE]; Muehlstrasse 9a, 82547
Eurasberg (DE). **OTTENWAEELDER**, Birgit [DE/DE];
Beinhofstrasse 1a, 81247 Muenchen (DE). **BAHR**,
André [DE/DE]; Raffaelweg 6, 40724 Hilden (DE).
DUESTERHOEFT, Andreas [DE/DE]; Karlrobert-Kre-
iten-Strasse 14, 40724 Hilden (DE). **KOENIG**, Christoph
[DE/US]; 6233 22nd Avenue N.E., Seattle, WA 98115
(US). **LAUBER**, Juergen [DE/DE]; Unterberg 1F, 42799
Leichlingen (DE). **HEUBNER**, Dagmar [DE/DE];
Grüne Trift 126a, 12557 Berlin (DE). **WAMBUTT**,
Rolf [DE/DE]; Florian-Geyer-Strasse 28, 12489 Berlin
(DE). **KOEHRER**, Karl [DE/DE]; Schlossmannstrasse
4, 40225 Duesseldorf (DE). **BEYER**, Andreas [DE/DE];

Helgolandring 106, 45149 Essen (DE). **GASSENHU-
BER**, Johann [DE/DE]; Emanuel Geibel Strasse 8,
65185 Wiesbaden (DE). **GRUBER**, Christian [DE/DE];
Zasinger Strasse 8, 81547 Muenchen (DE). **STRACK**,
Norman [DE/DE]; Linderbergweg 1, 82229 Seefeld
(DE). **MEWES**, H.W. [DE/DE]; Graf Toerring Strasse 9,
82237 Woerthsee (DE). **ANSORGE**, Wilhelm [DE/DE];
Boxberring 107/55, 69126 Heidelberg (DE). **GLASSL**,
Sabine [DE/DE]; Friedberger Weg 2, 64720 Michelstadt
(DE). **RITTMUELLER**, Claudia [DE/DE]; Siedler-
weg 2, 69151 Dilsbergerhof (DE). **REGIERT**, Thomas
[DE/DE]; Raiffeisenstrasse 38, 67227 Frankenthal
(DE). **BLOECKER**, Helmut [DE/DE]; Doeringstrasse
16, 38118 Braunschweig (DE). **BOECHER**, Michael
[DE/DE]; Alter Weg 41a, 38302 Wolfenbuettel (DE).
HORNISCHER, Klaus [DE/DE]; Mozartstrasse 2, 38106
Braunschweig (DE). **NORDSIEK**, Gabriele [DE/DE];
Ohfeld 34, 31188 Holle (DE). **TAMPE**, Jens [DE/DE];
Bergisch-Gladbacher-Strasse 656, 51067 Koeln (DE).

(74) Agents: **MERCER**, Christopher, Paul et al.; Carpmals
& Ransford, 43 Bloomsbury Square, London WC1A 2RA
(GB).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ,
DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,
HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,
LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,
NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,
TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG,
CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report

(88) Date of publication of the international search report:
20 June 2002

For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

(54) Title: HUMAN DNA SEQUENCES

(57) Abstract: Novel human cDNA sequence of a clones, the encoded protein sequence of a clones, antibodies and variants thereof, are provided. The disclosed sequence of a clones find application in a number of ways, including use in profiling assays. In this regard, various assemblages of nucleic acids or proteins are provided that are useful in providing large arrays of human material for implementing large-scale screening strategies. The disclosed sequence of a clones may also be used in formulating medicaments, treating various disorders and in certain diagnostic applications.



WO 01/12659 A3

INTERNATIONAL SEARCH REPORT

International Application No

PCT/IB 00/01496

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/12 C07K14/47 C12Q1/68 C07K16/18 A61K38/17
C12P21/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	WO 00 09552 A (GENETICS INST) 24 February 2000 (2000-02-24) Page 546, claim 86: SEQ.ID.No.: 77 ---	1-46
X	HILLIER L ET AL: "Human cDNA clone IMAGE:754267" EMBL SEQUENCE DATABASE, 23 July 1997 (1997-07-23), XP002163418 HEIDELBERG DE Accession Nr.: AA478899 abstract --- -/--	1-42



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search

20 March 2001

Date of mailing of the international search report

07.06.01

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl.
Fax: (+31-70) 340-3016

Authorized officer

De Kok, A

INTERNATIONAL SEARCH REPORT

International Application No
PCT/IB 00/01496

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HILLIER L ET AL.: "Human cDNA clone IMAGE: 754167" EMBL SEQUENCE DATABASE, 23 June 1997 (1997-06-23), XP002163419 HEIDELBERG DE Accession Nr.: AA478780 abstract	1-42
X	--- STRAUSBERG R ET AL.: "Human cDNA sequence IMAGE:2138166" EMBL SEQUENCE DATABASE, 24 March 1999 (1999-03-24), XP002163420 HEIDELBERG DE Accession Nr.:522149 abstract	1-42
X	--- HILLIER L ET AL.: "Human cDNA clone IMAGE:263887" EMBL SEQUENCE DATABASE, 5 January 1996 (1996-01-05), XP002163421 HEIDELBERG DE Accession Nr.: N28525 abstract	1-42
A	--- "Atlas(tm) human cDNA expression array I" CLONTECHNIQUES, April 1977 (1977-04), pages 4-7, XP002914393 US the whole document	1-20
A	--- REICHERT J ET AL: "HUMAN AND RODENT EXPRESSION PATTERN OF A FUSION GENE ISOLATED FROM AN MCF7 CDNA LIBRARY" INTERNATIONAL JOURNAL OF ONCOLOGY, vol. 9, no. 1, 1996, pages 29-32, XP000906725 page 29	1,6,7,17
A	--- WO 98 40486 A (GENETICS INST) 17 September 1998 (1998-09-17) page 29, line 20 -page 60, line 13 page 18, line 5 -page 26, line 32 -----	1-5, 8-25, 28-46

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IB 00/01496

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.: 21-40
because they relate to subject matter not required to be searched by this Authority, namely:
Rule 39.1(v) PCT - Presentation of information:
Although claims 21-40 could be considered as a mere presentation of information, according to Rule 39.1(v) PCT, the search has been carried out as far as possible in our systematic documentation.
2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-46 all partially

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

1. Claims: 1-46, all partially

Invention 1:

A nucleic acid molecule having the sequence of the clone hfbr2_16c16 (corresponding to SEQ.ID.1); an assemblage comprising said nucleic acid; a computer readable medium comprising said nucleic acid; a polypeptide encoded by said nucleic acid; an antibody binding to said polypeptide; an expression vector comprising said nucleic acid and a method for producing said polypeptide.

2. Claims: 1-46, all partially

Invention 2-233:

same as invention 1, but for each single clone as set forth in claim 1 (i.e. starting with clone hfbr2_16f21 and ending with clone hut1_2h3)

NB: for the sake of conciseness, the first subject-matter is explicitly defined, the other subject-matter by analogy thereto.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/IB 00/01496

Patent document cited in search report		Publication date		Patent family member(s)	Publication date
WO 0009552	A	24-02-2000	AU	5557099 A	06-03-2000

WO 9840486	A	17-09-1998	US	5976837 A	02-11-1999
			AU	6702298 A	29-09-1998
			EP	0973890 A	26-01-2000
